From Mice to Men

Klein Teeselink, Bouke

2021

Cover page

(This page is intentionally left blank)

From Mice to Men:

Field Studies in Behavioral Economics

ISBN: 978-90-361-0637-5

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 733 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

From Mice to Men:

Field Studies in Behavioral Economics

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de School of Business and Economics
op donderdag 11 maart 2021 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Bouke Klein Teeselink

geboren te Enschede

promotor:	prof.dr. M.J. van den Assem
copromotor:	dr. D. van Dolder
promotiecommissie:	prof.dr. T.J.S. Offerman
	prof.dr. J.J.M. Potters
	prof.dr. K.I.M. Rohde
	prof.dr. A.R. Soetevent
	prof.dr. G.U. Weitzel

Acknowledgments

I spent the past three years as an economist studying psychology at a finance department. The dissonance emanating from this eclectic combination of topics was oftentimes challenging, but at the same time contributed tremendously to my personal and academic growth, and eventually culminated in this dissertation. None of this would have been possible without my two supervisors, Martijn van den Assem and Dennie van Dolder. The fortunate decision to write my PhD under their supervision ironically set me off on a path of studying seemingly incorrect choices.

Like Sherlock and Watson, Tversky and Kahneman, and Simon and Garfunkel, Martijn and Dennie are one of those duos whose joint work is significantly better than the sum of their individual efforts. Their Popperian attitude towards their own research is inspiring as much as it is uncommon, and they both taught me valuable lessons about conducting and communicating research. I hope our collaboration will last a lifetime.

Martijn's impeccable attention to detail made my lack thereof all the more glaring, but it also encouraged to work on my deficiencies. In the same vein, I used to think that my writing was decent until the day Martijn commented on my first draft. Yet, Martijn's best quality is not his critical attitude *an sich*, but rather the fact that his comments, tips and criticism are invariably constructive, which contributed a lot to my academic development. If Martijn is ever to return a draft of mine uncommented, it will constitute the crown on my research career.

Like Martijn, Dennie is a researcher of great creativity, and his versatility in terms of skills and knowledge is inspiring. I have learned a lot from Dennie about structuring papers and placing findings in the existing literature. Furthermore, I can only appreciate his willingness to tolerate, or even engage with my myriad obscure deviations from the work I was actually supposed to do, which helped me to develop myself as an independent researcher.

Of course, I was also part of a bigger team in the Finance Department at the Vrije Universiteit Amsterdam. In spite of my often divergent research interests, I could not have wished for a more pleasant workplace than the one I experienced during the past three years. To my own surprise, my colleagues proved that finance academics need not be selfish, rude, arrogant and snobby, but can also be kind, collaborative and friendly instead. I am particularly grateful for my interactions with fellow PhD students Dieter, Susanne, Shihao, Ilka, Justus, Andries, Kirsten, Dries and Samet. Without our lunch conversations, I would have remained perfectly oblivious to many subfields of finance. I would also like to thank Debby Tielman both for her administrative support and for making the PhD-secretariat corner infinitely more lively.

I am furthermore grateful to Jason Dana, who hosted my research visit at Yale. Jason is a very warm person with an incredible depth and breadth of knowledge. He did a great job integrating me into the research group, and I benefited a lot from attending the weekly lab sessions, for which I would also like to thank, among others, Shane Frederick, Gal Zauberman, George Newman. Shane added an additional layer to my visit by shamelessly taking advantage of my inability to turn down exploring new and interesting data sets. I am also indebted to my other collaborators, Matthew Jordan, Nicholas Adolph and Rogier Potter van Loon. Without their contributions, some chapters of this dissertation would not exist.

At this point, I should mention my support system beyond the boundaries of my direct work environment. Akshath and Francis facilitated the necessary institution of afternoon drinks that helped me stay sane during my PhD, although I might have finished this dissertation a lot earlier without them. Similarly, my virtual rant sessions with Ioana provided me with an all-too-important frustration outlet throughout these years. I would also like to thank Derek, Remen, Laura, Rada, Peter, Sofia, Rossy and Lies for reminding me there exists a world outside academia.

A special note of gratitude goes to my family. On top of creating me, my parents also created an environment of freedom and responsibility that allowed me to explore my interests, to become my own person, and to develop myself intellectually. My brother and sister taught me how to collaborate at a very young age. Furthermore, without my late grandfather, who somehow managed to instigate a fascination for economics in a pubescent boy, my life might have taken a different path altogether.

Last but not least, I want to mention my partner Tanvi. She is my most important motivator. Every day she encourages me to be the best version of myself. She mitigates my worst impulses and strengthens my best. Without her, this journey would have been immeasurably lonely.

Contents

Chapter 1

Introduction

Behavioral economics principally emerged as a critique on the traditional assumption in economics that people have infinite cognition, intelligence, self-control and selfishness. Armed with lab experiments as its primary tool of investigation, behavioral economics endeavors to test and improve the behavioral foundations that underlie most important economic theories.

The findings of many such experiments have raised serious doubts about the appropriateness of rationality and selfishness axioms. For example, people appear to evaluate risky prospects not only by the absolute outcome, but also by the relative comparison to a certain reference point (Kahneman and Tverky, 1979; Tversky and Kahneman, 1992), they seem to care about the welfare of others (Fehr and Schmidt, 1999; Fehr and Gachter, 2000; Charness and Rabin, 2002), they oftentimes violate time-consistency (Thaler, 1981; Loewenstein and Prelec, 1992; Frederick et al., 2002) and when monetary incentives increase, their effort and performance sometimes deteriorate (Gneezy and Rustichini, 2000; Ariely et al., 2009).

In the light of these deviations, it may be surprising that behavioral economics is still not entirely accepted as part of the mainstream. This partial repudiation possibly arises from doubts about the external validity of laboratory studies (Levitt and List, 2007*a*,*b*, 2008). Critics argue that many behavioral anomalies would disappear in real-life situations where the experience and stakes are high and opportunities to learn are plentiful.

The current dissertation aims to explore whether these criticisms are justified. Rather than simply dismissing experimental evidence on the above-mentioned grounds, I use four naturally occurring data sets to test whether the principles that guide behavior in the laboratory can also explain the real-life decisions of athletes, game show participants, patients and darts players.

In Chapter 2, together with Martijn van den Assem, Dennie van Dolder and Jason Dana, I examine the optimality of strategic behavior in the *Showcase Showdown*, a high-stakes game in the American TV show *The Price is Right*. The *Showcase Showdown* is a simple sequential game with perfect information for which contestants can find the optimal strategy by backward induction. Studying contestants' decisions over a 40 year period, we show that they often deviate from the unique subgame perfect Nash equilibrium. We find that these deviations can neither be explained by random decision errors nor by a preference for harm caused by a failure to act over the equivalent harm caused by an explicit action. Instead, the observed behavior can be explained by limited foresight, where a contestant only thinks ahead to the next stage of the game.

In Chapter 3, Martijn van den Assem, Dennie van Dolder and I use professional sports matches to study the effect of marginally trailing on performance. Berger and Pope (2011) had previously shown that being slightly behind increases the likelihood of winning in professional and collegiate basketball. We extend their analysis to large samples of Australian football, American football and rugby matches, but find little to no evidence of such an effect for these three sports. When we revisit the phenomenon for basketball, we do find supportive evidence for National Basketball Association (NBA) matches from the period analyzed in Berger and Pope. However, we find no significant effect for NBA matches from outside this sample period, for collegiate matches, and for matches from the Women's NBA. High-powered meta-analyses across the different sports and competitions do not reject the null hypothesis of no effect of being slightly behind on winning.

In Chapter 4, Matthew Jordan, Nicholas Adolph, Shane Frederick and I explore the role of transaction utility in prescription medicine purchases. Pharmaceutical pricing in the United States is quite variable, and patients often do not know the price of their medications until they arrive at the pharmacy to retrieve their prescriptions. In recent years, pharmaceutical manufacturers have introduced discounts to reduce the out-of-pocket costs for patients. We exploit a unique data set containing transaction data from approximately

85 percent of all US pharmacies to estimate the causal effect of discounts on patients' propensity to purchase their medicine *at a given price*. We find an economically and statistically significant *mere discounts effect*: patients are more likely to buy their medicine at a given price if that price results from a discount. This shift of the demand curve is best explained by the transaction utility introduced by discounts.

In Chapter 5 of this dissertation, Rogier Potter van Loon, Martijn van den Assem, Dennie van Dolder and I examine how within-match variation in incentives affects the performance of darts players. The game of darts offers an attractive naturally occurring research setting, because performance can be observed at the individual level and without obscuring effects of risk considerations and behavior of others. We analyze four data sets covering a total of 29,381 darts matches of professional, amateur, and youth players. We find that amateur and youth players display a sizable performance decrease at decisive moments. Professional players appear less susceptible of such choking under pressure.

Each of these settings provides an appealing set of characteristics for studying behavioral phenomena in the field. As in most experiments, the decisions are of a simplicity that allows us to identify the determinants of people's behavior. Yet, compared to most experiments, the stakes are substantially higher. Incorrect decisions in *The Price is Right* and sports matches can result in large monetary or emotional losses, while prescription medicine purchases are potentially a matter of life and death. Moreover, people are highly familiar with the respective environment, either because they repeatedly perform the same action themselves (sports and medicine), or because they have normally observed many others who face similar choices (*The Price is Right*). Furthermore, even though athletes, game show participants and chronically ill patients may not be representative for the population in general, they are clearly different from the university students that typically participate in lab experiments. Because our current understanding of human decision making disproportionally relies on students, the informational content of studying new, dissimilar people is relatively high. If the same principles guide the behavior of students and darts players, for example, we can be reasonably sure that those principles form a robust component of human behavior.

Chapter 2

Backward Induction in the Field:

High-Stakes Evidence from The Price is Right[1]

2.1 Introduction

Many economic interactions are of a sequential nature. A negotiator who makes a bargaining offer, an entrepreneur who considers whether to enter a market and a director who decides how many goods to produce, all need to do so based on their best assessment of how their competitors will react (von Stackelberg, 1934; Selten, 1978; Dixit, 1982; Rubinstein, 1982). Such situations can often be modeled as sequential games of perfect information, for which the unique subgame perfect Nash equilibrium (USPNE) can be found through backward induction. Unfortunately, the descriptive accuracy of game theory is difficult to test in the field, because the predictions are sensitive to characteristics of the situation that are often not known: the set of options agents can choose from, as well as the information on which they base their choices. Therefore, most tests of the USPNE derive from lab experiments. Such experiments generally conclude that many people deviate from the USPNE, casting doubt on the descriptive validity of backward induction as a solution concept (Rosenthal, 1981; McKelvey and Palfrey, 1992; Fey et al., 1996; Binmore et al., 2002; Johnson et al., 2002; Levitt et al., 2011; Dufwenberg and Van Essen, 2018).

[1] This chapter is based on joint work with Martijn van den Assem, Dennie van Dolder and Jason Dana.

It is, however, still a topic of debate to what extent failures of backward induction in the lab can be extrapolated to the outside world. Critics argue that the external validity of experimental findings is compromised because the stakes are typically too low to adequately incentivize subjects to pursue the optimal strategy, and subjects are often unfamiliar with the task at hand (Binmore, 1999; Levitt and List, 2007*a*,*b*).[2] As a consequence of the limitations of field data, however, real-world evidence on the use of backward induction remains scarce.

The present chapter examines the optimality of strategic decisions in the *Showcase Showdown* (SCSD), a simple sequential game of perfect information in the long-running American TV show *The Price is Right*. In this game, three contestants take turns to once or twice spin a wheel that contains all multiples of 5 up to 100.[3] Each contestant's score is the sum of their first and, if chosen, second spin. The contestant whose total score is closest to 100 without exceeding it wins the game and proceeds to the *Showcase* round, where they compete with one other contestant to win a set of valuable prizes. Most spinning decisions involve a trade-off: spinning may increase a contestant's score, but at the same time it creates a risk of immediate defeat by exceeding 100 points. Coe and Butterworth (1995), Grosjean (1998) and Tenorio and Cason (2002) have derived the USPNE for this game. The equilibrium strategy takes the form of a decision rule that determines whether or not contestants should use their second spin.

The SCSD provides a number of characteristics that make it a particularly appealing setting to study game-theoretic predictions in the field. First, as in most experiments, there is no uncertainty about the potential choice options and information that contestants have. Second, the prospective prizes—on average worth tens of thousands of dollars—dwarf the incentives that are typically employed in experiments. Third, because the show has been on the air for a long time and participants self-select into attending the show

[2] See Falk and Heckman (2009); Camerer (2015) for a critical reply.

[3] Throughout the rest of the chapter, we refer to the contestant who spins first as 'Contestant 1', to the contestant who spins second as 'Contestant 2', and to the contestant who spins last as 'Contestant 3'.

and possibly being a contestant, they can be expected to be highly familiar with the rules and setup of the game.[4]

We study a large sample of 6,179 renditions of the *Showcase Showdown*, covering the decisions of 18,537 contestants. Although this sample may seem sizable, almost all Contestant 3's decisions are trivial and indeed conform to the following rule: always spin again if her first spin is lower than the best preceding score, and always stop if it is higher. We therefore focus all our analyses on the decisions of Contestant 1 and 2.[5] Our first analysis examines whether contestants adhere to the USPNE. Interestingly, we find that Contestant 1 makes more errors than Contestant 2, and that Contestant 1 almost exclusively errs by underspinning. Contestant 2's mistakes, by contrast, are roughly equally distributed between underspinning and overspinning.[6]

We consider several explanations for the observed deviations from the USPNE. First, we examine whether people depart from the equilibrium strategy because they make random decision errors. If a contestant's expected costs of underspinning are lower than her expected costs of overspinning, a decision error of a given magnitude would more often result in the former than in the latter, and may thus explain the observed underspinning pattern of Contestant 1.

A second reason why contestants may deviate from the USPNE is that they are possibly prone to omission bias—a preference for harm caused by a failure to act over the equivalent harm caused by an explicit action (Ritov and Baron, 1990; Spranca et al., 1991). Because decisions in the SCSD posit a clear choice between action (spinning) and inaction (not spinning), omission bias provides a preference-based explanation for why contestants may underspin compared to the equilibrium strategy. Prior research shows

[4]Game shows have previously been used to study a wide range of topics in economics, such as decision making under risk (Gertner, 1993; Metrick, 1995; Post et al., 2008), discrimination (Levitt, 2004; Belot et al., 2010), strategic reasoning (Bennett and Hickman, 1993; Berk et al., 1996), bargaining (van Dolder et al., 2015), and cooperation (List, 2006; Oberholzer-Gee et al., 2010; van den Assem et al., 2012; Turmunkh et al., 2019).

[5]Contestant 1 and 2 also face relatively many easy choices. A contestant whose first spin is very low or, in the case of Contestant 2, lower than Contestant 1's score, should always spin again. A contestant whose first spin is close to 100 should obviously stop spinning. Our large sample ensures that we have sufficiently many non-trivial decisions.

[6]Tenorio and Cason (2002) find a similar pattern of behavior both in a relatively small sample of Price is Right episodes, as well as in a laboratory experiment. They argue that this finding can be explained by omission bias—a preference for harm caused by a failure to act over the equivalent harm caused by an explicit action.

that omission bias plays an important role in settings where decision makers face a similar choice between action and inaction such as blackjack and sports refereeing (Keren and Wagenaar, 1985; Carlin and Robinson, 2009; Moskowitz and Wertheim, 2011).[7]

A third possible explanation for deviations from the USPNE is that contestants adopt a simplified representation of the decision problem. Economists have proposed the idea that people have limited foresight and look at most a few steps ahead in strategic situations with multiple stages (Jehiel, 1995, 1998; Gabaix and Laibson, 2005; Gabaix et al., 2006; Mantovani, 2015; Rampal, 2018; Ke, 2019). In our implementation of limited foresight, we allow for the possibility that a fraction of spinning choices is based on the next stage of the game only. In practice, this means that Contestant 1 only tries to beat Contestant 2 and ignores the presence of a third contestant. Because Contestant 2 only needs to look one step ahead in the first place, limited foresight corresponds to backward induction for her.

We formulate our models of strategic decision making as Agent Quantal Response Equilibria (AQRE; McKelvey and Palfrey, 1998). The quantal response equilibrium is a stochastic generalization of the Nash equilibrium. The general idea is that agents make random miscalculations in evaluating the utility of each outcome, and anticipate that their competitors make similar decision errors. As a result, the model defines a probabilistic choice function, where better responses are more likely to be chosen. Quantal response equilibria have been used to successfully explain behavior across a multitude of experimental games (see for example Capra et al., 1999; Anderson et al., 2001; Goeree et al., 2002, 2003; Cai and Wang, 2006). Applied to our setting, the AQRE predicts the likelihood that a contestant will use their second spin. This likelihood increases in the relative attractiveness of spinning compared to not spinning. To test the alternative explanations for deviating from the USPNE, we also estimate alternative structural models in which we incorporate omission bias and limited foresight.

To evaluate the descriptive validity of these explanations, we compare each model's predicted spinning likelihood to the actual fraction of contestants that use their second

[7]A related psychological phenomenon to omission bias is sudden death aversion: the tendency to avoid strategies that have a higher chance of final success but include the possibility of immediate defeat in favor of lower-success strategies without the possibility of losing directly (Walker et al., 2018). Sudden death aversion leads to the same prediction as omission bias in our setting.

spin. The results show that the general pattern of behavior cannot be explained either by decision errors alone, or by a combination of decision errors and omission bias. Although both models describe contestants' spinning decisions more accurately than the USPNE, substantial discrepancies between observed and predicted behavior remain. Instead, the observed choices can be explained well by limited foresight. Our results suggest that approximately 40 percent of the contestants simplify the game by looking only one step ahead.

Because *The Price is Right* has been running for more than 45 years, a relevant question is whether contestants have become progressively more likely to follow the equilibrium strategy. To explore such a learning effect, we split our sample in four and estimate the limited foresight model for each of these time periods. The results show that the proportion of spinning decisions that are consistent with backward induction increases from 55 percent between 1979 and 2001 to 68 percent between 2014 and 2018. Yet, despite this learning effect, many contestants still do not follow the equilibrium strategy, even after several decades of *The Price is Right*.

The remainder of the chapter is structured as follows. Section 2.2 describes the the game show in more detail. Section 2.3 outlines the equilibrium strategies for each player. Section 2.4 provides an overview of the data and a preliminary analysis of equilibrium play. Section 2.5 describes the structural models. Section 2.6 shows the main results. Section 2.7 provides several robustness checks. Section 2.8 shows an analysis of learning. Section 2.9 concludes and discusses our findings.

2.2 Game Description

The Price is Right is a television game show that aired in 1972 and has been running uninterruptedly ever since. Each episode contains six auction games and six pricing games. For each of the auctions, four contestants are selected from the studio audience to participate in a game where they have to guess the price of a retail consumer product such as a microwave or a TV. The contestant whose guess is closest to, but not higher than, the retail price of the product wins and proceeds to the *Showcase Showdown* (SCSD).

Before participating in the SCSD, each auction winner plays a pricing game where she can win several prices. There are currently 77 different pricing games in rotation.

The SCSD is our main focus. In this game, played twice every episode, three participants take turns to spin a wheel that contains all multiples of 5 up to 100. Ranked from lowest to highest in terms of prior winnings, each contestant spins the wheel once or twice. Their score is the outcome of the first spin if they spin once, and the sum of the two spins if they spin twice. The contestant whose total score is closest to 100 without exceeding it wins the game and proceeds to the *Showcase* round. If multiple contestants tie for the highest score, they enter a spin-off in which each tied contestant spins the wheel once more, and the one who scores the highest number of points wins. In the case of further ties, this procedure is repeated until a winner emerges. If contestants score exactly 100 points, they obtain a bonus prize of $1,000 plus an additional bonus spin that either yields $10,000 if the contestant scores 5 or 15 points, or $25,000 if she scores 100 points again. If multiple contestants tie at a score of 100, the score of their bonus spin counts as their spin-off score.

The winners of both renditions of the SCSD proceed to the *Showcase* round. In this final game, each of the two contestants guesses the retail value of their own respective showcase, which consists of various non-monetary prizes such as cars and trips. The contestant whose guess is closest to the retail price without exceeding it wins the content of her showcase. If the winner's guess is within a specified amount of the retail price, she wins both showcases, and if both guesses exceed the retail price, the showcases remain unclaimed.

2.3 Unique Subgame Perfect Nash Equilibrium

Various researchers have derived the USPNE for the *Showcase Showdown* under the assumption that contestants are risk neutral (Coe and Butterworth, 1995; Grosjean, 1998; Tenorio and Cason, 2002). Following the notation of Tenorio and Cason (2002), let a_i and b_i denote the points obtained in the first and, if chosen, second spin by contestant i, where

$i \in \{1, 2, 3\}$. We assume $a_i, b_i \sim$ Discrete Uniform$[5, 10, \ldots, 100]$.[8] Let $t_i = a_i + b_i$ denote contestant i's final score, and let x_i be the lowest score with which she has a chance to win. While Contestant 1 can theoretically win the SCSD with any first spin, Contestant 2 and 3 must at least equalize the best preceding score to have a positive winning chance.

Let E_s denote the expected monetary value of winning the *Showcase* round. Let E_{b_1} the cash bonus from obtaining a score of 100, and let E_{b_2} and E_{b_3} be the additional cash bonuses obtained from spinning either 5 or 15 (E_{b_2}) or 100 (E_{b_3}) in the bonus spin. For simplicity, we assume that contestants' perceived chance of winning the *Showcase* round after winning the SCSD is equal to 50 percent.[9] We furthermore assume that contestants' subjective belief about the value of the showcase is equal to the average retail price of all showcases in the previous five years.[10]

To illustrate the optimal strategy, consider the reward scheme of 2018, when the expected showcase value was equal to \$28,965 and the bonus prizes were equal to \$1,000, \$10,000 and \$25,000. Independent of the reward scheme, Contestant 3's optimal strategy is relatively straightforward: always spin again if $a_3 < x_3$ and always stop spinning if $a_3 > x_3$. Given the relative simplicity of this strategy, the host generally does not even present the contestant with a choice and simply takes the optimal choice as given. She only faces a non-trivial decision in relatively rare occasion that she ties the best preceding score ($a_3 = x_3$). In this case, she should spin again if her first spin is 50 points or fewer ($t_1 \neq t_2$) or 65 points or fewer ($t_1 = t_2$).

Contestant 2 faces a more complex decision, because she has to anticipate the behavior of Contestant 3. With the exception of trivial choices in which her first spin is below the score of Contestant 1 ($a_2 < x_2$), Contestant 2 needs to balance the benefit of getting a better score to beat Contestant 3 against the risk of self-elimination from exceeding 100

[8]In theory, contestants may have some degree of control over the wheel. There are, however, several reasons why the assumption of uniformly distributed scores seems reasonable. First, contestants are unlikely to have much prior experience with spinning a huge wheel. Second, contestants' ability to spin any specific number is severely compromised by the fact that the wheel needs to make at least one full rotation for a spin to be valid, and the numbers on the wheel are quasi-randomly distributed.

[9]The empirical winning probability is slightly lower, because neither contestant wins the showcase if both guess a price that is higher than the retail price of their respective showcases. This is partially offset by the fact that contestants win *both* showcases if their guess is within a specified amount of the actual retail price. We examine the sensitivity of our results to this assumption in Section 2.7.1.

[10]Our results do not materially change if contestants assume that the value of the showcase is equal to the average prize of the current year.

points. Given Contestant 3's equilibrium behavior, Contestant 2 maximizes her expected gains when she spins again for first spins of 55 or fewer points that beat the score of Contestant 1. If her first spin ties the score of Contestant 1's, she should spin again for first spins of 65 points or fewer.

Contestant 1 can determine her optimal strategy by reasoning backwards through her opponents' equilibrium strategies. This yields an optimal spinning threshold of 65 that determines she should spin again for first spins that are equal to or lower, and stop otherwise.

The USPNE can vary slightly depending on the expected showcase value compared to the prevailing bonus scheme. Proposition 2.1 and 2.2 outline the USPNE per contestant both for Bonus Scheme 1 (1979 to mid-2008, $E_{b1} = \$1,000$, $E_{b2} = \$5,000$ and $E_{b3} = \$10,000$) and for Bonus Scheme 2 (mid-2008 to 2018, $E_{b1} = \$1,000$, $E_{b2} = \$10,000$ and $E_{b3} = \$25,000$).[11]

Proposition 2.1. *Under Bonus Scheme 1, the USPNE is as follows:*

- *Contestant 1 should spin again if she gets 70 points or fewer when* $E_s < \$821$*, and 65 points or fewer when* $E_s \geq \$821$*.*

- *Contestant 2 should spin again if she beats Contestant 1 with a score of 60 points or fewer when* $E_s < \$2,564$*, 55 points or fewer when* E_s *is between* $\$2,564$ *and* $\$27,826$*, and 50 points or fewer when* $E_s > \$27,826$*. She should also spin again if she equalizes Contestant 1 with a score of 70 points or fewer when* $E_s < \$10,702$*, and 65 points or fewer when* $E_s \geq \$10,702$*.*

- *Contestant 3 should spin again if she beats Contestant 1 and 2 with any first spin when* $E_s > \$3,900$*. She should furthermore spin again if she equalizes one other contestant with a score of 55 points or fewer when* $E_s < \$2,900$*, and 50 points or fewer when* $E_s \geq \$\$2,900$*. She should also spin again if she equalizes both other contestants with a score of 70 points or fewer when* $E_s < \$3,900$*, and 65 points or fewer when* $E_s \geq \$3,900$*.*

Proposition 2.2. *Under Bonus Scheme 2, the USPNE is as follows:*

[11] We examine the possibility that contestants ignore the bonus prizes altogether in Section 2.7.2.

- *Contestant 1 should spin again if she gets 70 points or fewer when* $E_s < \$1,334$*, and 65 points or fewer when* $E_s \geq \$1,334$*.*

- *Contestant 2 should spin again if she beats Contestant 1 with a score of 60 points or fewer when* $E_s < \$4,167$*, 55 points or fewer when* E_s *is between* $\$4,167$ *and* $\$45,217$*, and 50 points or fewer when* $E_s > \$45,217$*. She should also spin again if she equalizes Contestant 1 with a score of 70 points or fewer when* $E_s < \$17,391$*, and 65 points or fewer when* $E_s \geq \$17,391$*.*

- *Contestant 3 should spin again if she beats Contestant 1 and 2 with any first spin when* $E_s > \$6,338$*. She should furthermore spin again if she equalizes one other contestant with a score of 55 points or fewer when* $E_s < \$4,713$*, and with 50 points or fewer when* $E_s \geq \$4,713$*. She should also spin again if she equalizes both other contestants with a score of 70 points or fewer when* $E_s < \$6,338$*, and 65 points or fewer when* $E_s \geq \$6,338$*.*

2.4 Data Description and Preliminary Results

2.4.1 Data Description

We scraped data from *The Price is Right Episode Guide*, a fan-edited forum that maintains detailed recaps of *The Price is Right* episodes, on 10 August 2018. For each rendition of the SCSD, we obtained the scores of the first and, if chosen, second spin for each contestant. When possible, we also obtained the retail value of both showcases displayed in the *Showcase* round. We exclude forum pages that have a different format than the majority of the pages and could therefore not be straightforwardly scraped, as well as SCSDs that contain obvious mistakes such as scores that are not from the set $5, 10, \ldots, 100$. We also exclude all special episodes in which the bonus prizes deviate from the prevailing bonus scheme. As a result of imposing these restrictions, our final sample contains 6,178 of the available 10,930 renditions of the SCSD, covering 3,403 episodes that took place between 1979 and 2018. The forum itself serves as a quality check on the data because attentive forum members identify and correct most mistakes. Comparing several episode recordings to our final data set revealed no further inaccuracies.

Table 2.1 presents summary statistics for four different time periods: (i) from 1 January 1979 to 31 December 2001, (ii) from 1 January 2002 to 21 September 2008, (iii) from 22 September 2008 to 31 December 2013, and (iv) from 1 January 2014 to 4 July 2018. On 22 September 2008, the bonus scheme changed from Bonus Scheme 1 ($E_{b1} = \$1,000$, $E_{b2} = \$5,000$ and $E_{b3} = \$10,000$) to Bonus Scheme 2 ($E_{b1} = \$1,000$, $E_{b2} = \$10,000$ and $E_{b3} = \$25,000$).

The average value of the showcase ranges from \$19,425 in Period 1 to \$29,345 in Period 3. Proposition 2.1 and 2.2 show that these prizes ensure that the optimal spinning threshold for Contestant 1 is 65 for the complete sample. For Contestant 2, the threshold when she beats Contestant 1 with her first spin is 55 in Period 1, 3 and 4, and 50 in Period 2.[12]

Contestant 1, 2 and 3 use their second spin 60, 70 and 76 percent of the time. Contestant 1's spinning rate gradually increases over time, whereas the spinning rates of both Contestant 2 and 3 remain relatively stable.

If all contestants follow the equilibrium strategy in Period 1, 3 or 4, Contestant 1 wins the SCSD 30.9 percent of the time, Contestant 2 wins 32.9 percent of the time and Contestant 3 wins 36.2 percent of the time. In Period 2, the equilibrium winning chances are 30.8, 33.0 and 36.2 percent. In practice, Contestant 1 wins 30 percent of all SCSDs, whereas Contestant 2 and 3 win 33 and 37 percent of the time.

2.4.2 Preliminary Results

As a first analysis, we explore the extent to which contestants' spinning decisions are consistent with the USPNE. Because almost all Contestant 3's choices are trivial, we exclude them from the analysis. For the same reason, we exclude Contestant 2's decisions that follow first spins that are below the final score of Contestant 1. For the current analysis, we furthermore exclude Contestant 2's choices that either follow ties with Contestant 1 or take place between 2002 and 2008, because the USPNE is different in both cases.

Figure 2.1 displays the proportion of spinning choices that deviate from the USPNE for Contestant 1 (Panel A) and Contestant 2 (Panel B). The dark grey bars show the fraction

[12] Table 2.7 in the Appendix gives an overview of the optimal spinning thresholds per year.

Table 2.1: Summary statistics

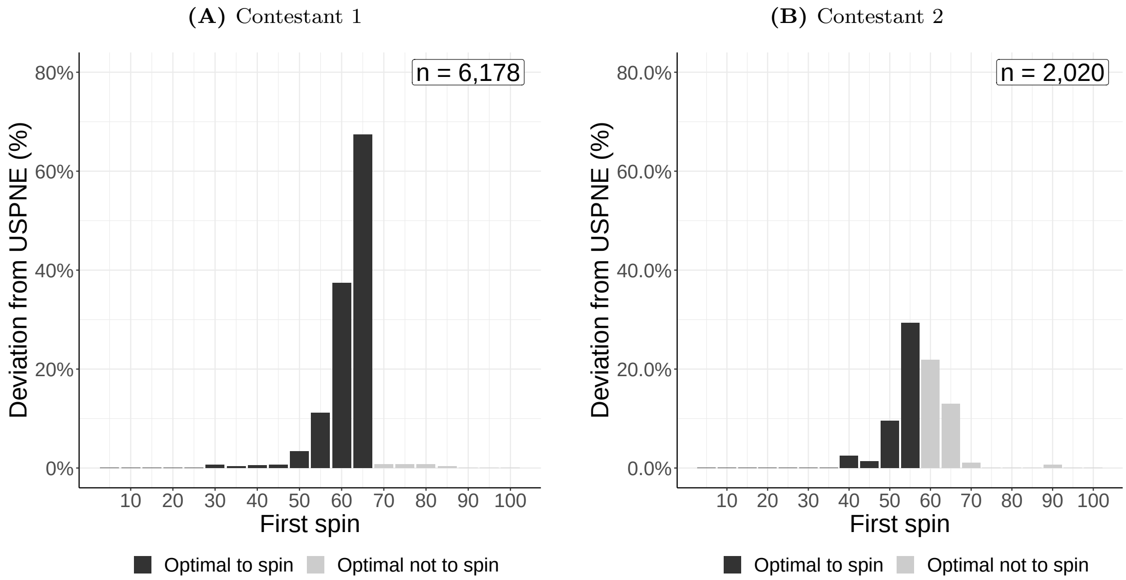
	Full sample	1979 to 2001	2002 to mid-2008	Mid-2008 to 2013	2014 to 2018
Panel A: General stats					
Episodes	3,403	935	945	806	717
SCSDs	6,178	1,672	1,772	1,439	1,295
Bonus scheme	1-2	1	1	2	2
Panel B: Showcase value					
Average	\$26,005	\$19,425	\$27,687	\$29,345	\$27,725
Std. dev.	\$9,706	\$8,829	\$12,332	\$7,217	\$7,122
Minimum	\$760	\$760	\$13,322	\$17,954	\$3,444
Median	\$25,938	\$18,634	\$25,988	\$28,128	\$27,418
Maximum	\$174,018	\$52,529	\$174,018	\$63,035	\$60,676
Panel C: Spinning statistics					
C1 spins	0.604	0.587	0.590	0.612	0.639
C2 spins	0.698	0.696	0.686	0.698	0.717
C3 spins	0.756	0.750	0.758	0.769	0.747
Panel D: Winning fractions					
C1 wins	0.299	0.294	0.306	0.297	0.299
C2 wins	0.335	0.340	0.328	0.345	0.328
C3 wins	0.366	0.366	0.366	0.358	0.373

Notes: The table displays the summary statistics for four different time periods. Panel A shows the number of episodes (*Episodes*) and renditions of the *Showcase Showdown* (*SCSD*) in our sample. *Bonus scheme* refers to the prevailing bonus scheme. In Bonus Scheme 1, the bonus prizes are equal to \$1,000, \$5,000 and \$10,000, and in Bonus Scheme 2, they equal \$1,000, \$10,000 and \$25,000. Panel B shows the average, standard deviation, minimum, median and maximum value of the showcase prize. In Panel C, *C1*, *C2* and *C3 spins* indicate the proportions of Contestant 1, 2 and 3 that use their second spin. *C1*, *C2* and *C3 wins* in Panel D give the fractions of SCSDs that are won by Contestant 1, 2 and 3.

of errors in situations where it is optimal to spin, and the light grey bars show the fraction of errors in situations when it is not. Panel A shows that Contestant 1 frequently departs from the equilibrium strategy if she scores between 55 and 65 points with her first spin. Because her optimal spinning threshold is 65, these deviations indicate that Contestant 1 underspins compared to the theoretical prediction.

Panel B shows that Contestant 2 departs considerably less frequently than Contestant 1. Even for first spins of 55—the score with the highest proportion of deviations—70 percent of contestants correctly opt for a second spin. Moreover, in contrast to Contestant 1, the errors are approximately symmetrically distributed between underspinning and overspinning.

Figure 2.1: Deviations from the USPNE



Notes: The figure shows the fraction of decisions that deviate from the USPNE for Contestant 1 (Panel A) and Contestant 2 (Panel B). For Contestant 2, we only consider non-trivial decisions in which her first spin beats the score of Contestant 1, and we exclude decisions between 2002 and 2008, because her optimal spinning threshold during this period is 50 instead of 55.

2.5 Structural Models

2.5.1 AQRE

Inherent in the USPNE is the assumption that all contestants flawlessly follow the equilibrium strategy. In reality, however, a certain degree of errors can logically be expected. We therefore examine whether contestants' deviations from the USPNE can be explained by random decision errors. In particular, if Contestant 1's costs of underspinning are lower than her costs of overspinning, decision errors may account not only for the frequency of the deviations, but also for their sign. Table 2.6 indeed indicates the presence of such a cost asymmetry in 2018.[13]

To formally explore whether decision errors can reconcile contestants' choices with the equilibrium predictions, we estimate an Agent Quantal Response Equilibrium (AQRE). The main idea underlying the AQRE is that players make random errors in calculating the utility associated with each outcome. Let $j \in 1, 2, \ldots J_i$ be the set of contestants that spin at position i, $i \in \{1, 2, 3\}$. Let R be the prevailing reward scheme. For risk neutral

[13] A similar asymmetry applies to the rest of our sample.

Contestant ij with first spin $a_{ij} = a$ and minimum score-to-beat $x_{ij} = x$, the expected utility $U_{ij}(s, x, a, R)$ from action $s \in \{\text{Spin}, \text{Not spin}\}$ corresponds to the expected prize money following that action. Because she makes decision errors, however, she mistakenly evaluates the expected utility of each strategy by random utility $\hat{U}_{ij}(s, x, a, R)$:

$$\hat{U}_{ij}(s, x, a, R) = U_{ij}(s, x, a, R) + \varepsilon_{ij} \tag{2.1}$$

The payoff disturbances ε_{ij} are distributed according to a density function $f(\varepsilon)$, with $E[\varepsilon_{ij}] = 0$. A contestant decides to spin if $\hat{U}_{ij}(s = \text{Spin}, x, a, R) \geq \hat{U}_{ij}(s = \text{Not spin}, x, a, R)$ and to stop otherwise. U and f together induce a predicted choice distribution over all observed spinning decisions. To obtain these predictions, let $S_{ij(U)}$ be the set of realizations of ε such that spinning yields a higher perceived payoff than not spinning for a given value of U. The predicted likelihood that contestant ij uses her second spin is then given by

$$P_{ij}(U) = \int_{S_{ij(U)}} f(\varepsilon)\mathrm{d}\varepsilon \tag{2.2}$$

Following convention, we assume that ε_{ij} are independently and identically distributed according to an extreme value distribution, which leads to choice probabilities following a logit distribution:

$$P_{ij}(U) = \frac{e^{\lambda_i U_{ij}(s=\text{Spin},x,a,R)}}{e^{\lambda_i U_{ij}(s=\text{Spin},x,a,R)} + e^{\lambda_i U_{ij}(s=\text{Not spin},x,a,R)}} \tag{2.3}$$

λ_i can be interpreted as the rationality parameter or payoff sensitivity of contestant i. If $\lambda_i \to 0$, contestants always spin with a 50 percent chance and if $\lambda_i \to \infty$, contestants choose the optimal strategy with certainty.

An important feature of the AQRE is that players have statistically accurate beliefs, meaning that each contestant expects the other contestants to behave in line with the model's predictions. For example, if the estimated likelihood that Contestant 2 uses her second spin is 70 percent when her first spin is 50, Contestant 1 correspondingly makes this assumption in her expectation of Contestant 2's behavior. Implicit in this argument are the assumptions that contestants take into account each other's mistakes, and use backward induction to devise their strategies.

We use maximum likelihood estimation (MLE) to estimate the model's parameters. We focus our analysis on all non-trivial choices of the first two contestants, including situations in which Contestant 2 ties Contestant 1 with her first spin. To avoid overfitting the model to the decisions by Contestant 1, we weight all observations to give the same aggregate decisions weights to Contestant 1 and 2.[14] We furthermore convert all prizes into 2015 dollars using the US Consumer Price Index (OECD, 2020).

2.5.2 Omission Bias

Our second model examines whether the deviations from the USPNE can be explained by omission bias. To formally test this hypothesis, we maintain the assumption that contestants make decision errors and additionally augment the risk-neutral utility function with an additional non-pecuniary component regarding the way L in which they may lose, where $L \in \{\text{Self-elimination}, \text{Elimination by another contestant}\}$. We denote the utility of self-elimination by γ and fix the utility from being eliminated by another contestant at 0.

Since omission bias presupposes that contestants have a stronger dislike for losing through an explicit action than through a failure to act, the parameter γ should take a negative value. We embed this augmented preference in the AQRE model. Hence, we implicitly assume that contestant correctly anticipate the influence of omission bias on their competitors' decisions.

2.5.3 Limited Foresight

In our third model, we consider that contestants may deviate from the USPNE because they do not (correctly) apply backward induction. Instead, we assume that a fraction of contestants makes choices consistent with limited foresight by only considering the next contestant in their spinning decision. Looking only one step ahead distorts Contestant 1's incentive to spin, because she will ignore the fact that a higher score will help her not only to beat Contestant 2, but also to defeat Contestant 3. Because Contestant 2

[14] The results do not materially change when we run the analyses without weighting.

only has to look one step ahead in the first place, her predicted behavior under limited foresight corresponds to backward induction.

To formally estimate the limited foresight model, we assume that a fraction β of Contestant 1's choices is consistent with limited foresight, whereas the remaining fraction $1 - \beta$ is consistent with correctly applying backward induction. We assume that when Contestant 1 bases her decision on a limited foresight horizon, she also expects Contestant 2 to ignore the presence of Contestant 3. This means that Contestant 1 expects that Contestant 2 always stops spinning if she beats Contestant 1's score with her first spin.[15] We maintain the assumption that contestants make random mistakes in their utility calculations, although it should be noted that the limited foresight model is not technically an equilibrium model because contestants' beliefs may not be consistent with the model's prediction.

To examine whether a preference for harm caused by omission over harm caused by commission can explain contestants' decisions beyond what is already explained by the limited foresight account, we additionally estimate a limited foresight model that incorporates omission bias.

2.6 Estimation Results

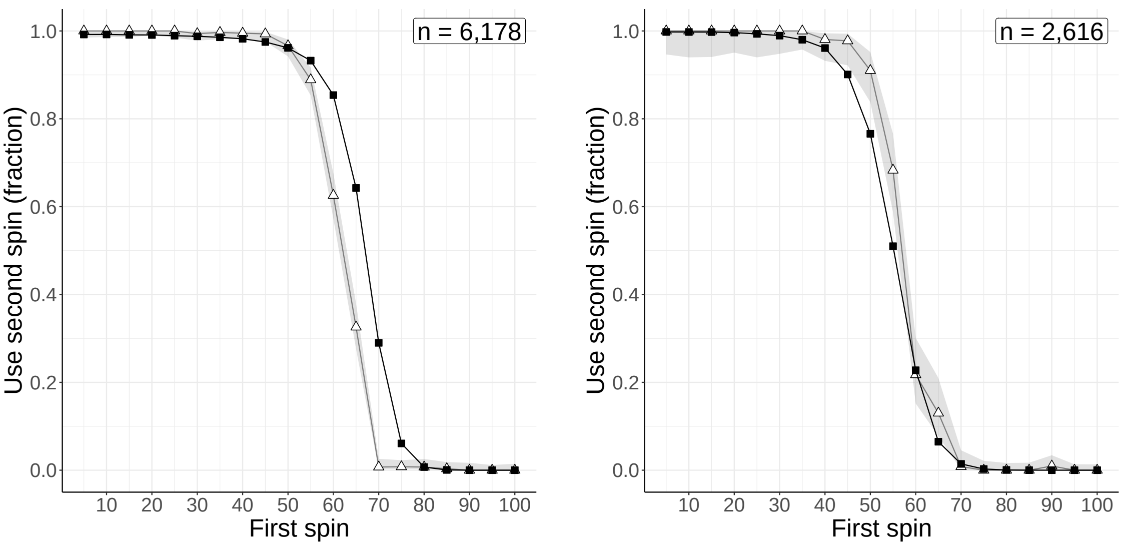
2.6.1 Baseline Model

In our baseline analysis, we estimate an AQRE model in which contestants are assumed to be risk neutral. Table 2.2 gives the parameter estimates. To visually examine the model's descriptive performance, we compare the model's predicted spinning probabilities to the actual fraction of contestants that use their second spin. Figure 2.2, Panel A shows a clear discrepancy between Contestant 1's actual behavior and the prediction of the AQRE model. When faced with difficult decisions—first spins between 50 and 70—Contestant 1 is considerably less inclined to spin than predicted. At the same time, Panel B indicates that the spinning decisions of Contestant 2 are roughly consistent with the equilibrium predictions, suggesting that random decision errors can explain most of her deviations.

[15] Assuming that Contestant 1's limited foresight prediction of Contestant 2's behavior is equal to Contestant 2's empirically observed spinning distribution does not materially change our results.

Figure 2.2: Observed and predicted spinning rates for the baseline AQRE model

(A) Contestant 1 **(B)** Contestant 2



Notes: The figure shows the actual spinning rate and the prediction of the baseline AQRE model, both for Contestant 1 (Panel A) and Contestant 2 (Panel B). The black curve with black squares shows the model's prediction, and the grey curve with white triangles depicts the actual behavior. The shaded areas around the curves with the actual behavior represent the 95 percent confidence intervals.

To quantitatively examine the descriptive accuracy of the baseline model, we first calculate the hit rate for both contestants. A decision is defined as a 'hit' if the model assigned at least a 50 percent probability to its occurrence. Following this definition, the baseline model correctly predicts 93.4 percent of Contestant 1's decisions and 96.5 of Contestant 2's. Although this might seem high, it should be noted that most decisions are relatively easy. For difficult decisions following first spins between 50 and 70, the hit rate decreases to 76.4 percent for Contestant 1 and to 84.0 percent for Contestant 2. Although the hit rate of the baseline AQRE model is identical to that of the USPNE in Table 2.2, this need not indicate that both models explain choices equally well, because the binary nature of a hit does not include important information about *how* correct a prediction is.

The hit rate is also uninformative about the extent to which the model under- or overpredicts spinning choices. To get an insight into the sign of the deviations, we calculate the overspinning rate. The overspinning rate is the average difference between contestants' actual spinning decisions, which take a value of either 0 or 1, and the corresponding predictions, which can take any value between 0 and 1. We find that Contestant 1

Table 2.2: Estimation results and goodness-of-fit structural models

	USPNE		Baseline		Omission bias		Limited foresight		OB & LF	
λ_1	-	-	0.0014	(0.0000)	0.0018	(0.0000)	0.0016	(0.0001)	0.0016	(0.0000)
λ_2	-	-	0.0011	(0.0000)	0.0010	(0.0000)	0.0011	(0.0000)	0.0011	(0.0000)
γ	-	-	-	-	-1056.0	(2.1)	-	-	173.7	(3.0)
β	-	-	-	-	-	-	0.410	(0.020)	0.457	(0.029)
N	8,943		8,943		8,943		8,943		8,943	
Log likelihood	-		$-1{,}679$		$-1{,}547$		$-1{,}455$		$-1{,}453$	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
Hit rate	0.934	0.965	0.934	0.965	0.952	0.953	0.952	0.965	0.952	0.967
Hit rate (50-70)	0.764	0.840	0.764	0.840	0.830	0.782	0.830	0.840	0.830	0.851
Overspinning	-0.063	0.001	-0.042	0.019	-0.013	0.040	0.001	0.019	0.001	0.016
Overspinning (50-70)	-0.234	0.045	-0.162	0.068	-0.055	0.143	-0.002	0.068	-0.002	0.054

Notes: The table shows the results and the goodness-of-fit for various models of strategic decision making. *USPNE* is the unique subgame perfect Nash equilibrium, *Baseline* is the baseline AQRE model, *Omission bias* is the AQRE model that incorporates omission bias, *Limited foresight* is the model in which a fraction of decisions is consistent with limited foresight, and *OB & LF* is the limited foresight model that incorporates omission bias. λ_1 and λ_2 are the rationality parameters of Contestant 1 and 2 respectively. γ expresses the utility difference between self-eliminating and being eliminated by another contestant. β is the estimated fraction of decisions that accord to limited foresight. The standard errors are given in parentheses. *Log likelihood* is the log likelihood value of the estimation, and N is the total number of non-trivial spinning decisions by Contestant 1 and 2 combined. *Hit rate* and *Hit rate (50-70)* are the fractions of correctly predicted decisions for all decisions and for decisions that follow first spins between 50 and 70. *Overspinning* and *Overspinning (50-70)* show the average difference between the actual and predicted spinning rates for all decisions and for decisions that follow first spins between 50 and 70. The goodness-of-fit measures are given separately for Contestant 1 (*C1)* and Contestant 2 (*C2*).

underspins compared to the prediction of the baseline model. The average difference is -4.2 percentage points for all choices combined, and -16.2 percentage points for first spins between 50 and 70. Although the difference is substantial, it constitutes a considerable improvement compared to the fit of the USPNE. Contestant 2's overspinning rate is lower than Contestant 1's. Across all choices, Contestant 2 is 1.9 percentage points more likely to spin than predicted, and for first spins between 50 and 70, this difference is 6.8 percentage points.

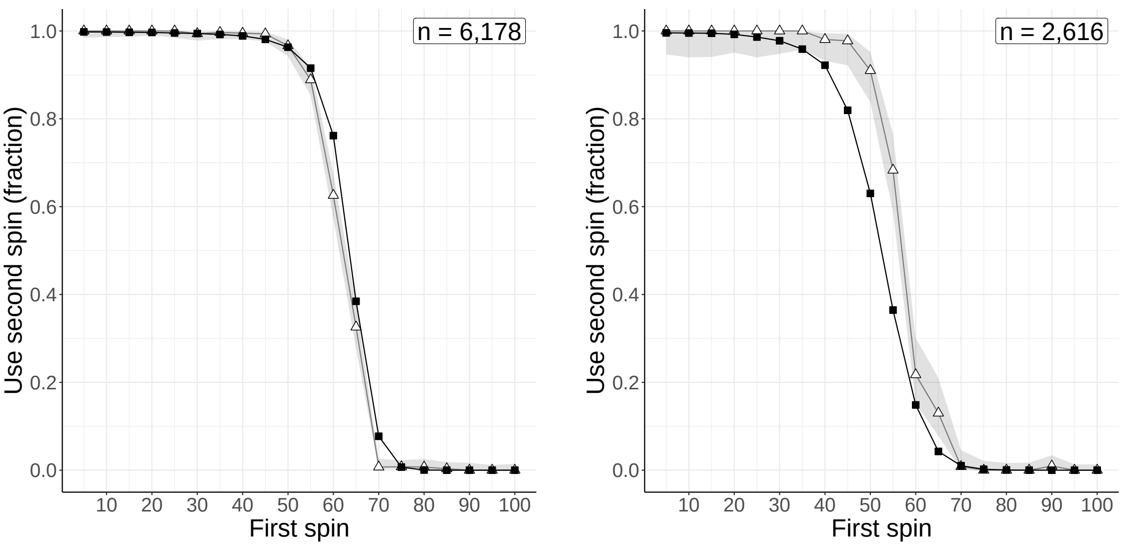
2.6.2 Omission Bias

In our second analysis, we extend the AQRE model with omission bias. The results in Table 2.2 show an estimated parameter γ of $-1{,}056.0$. This coefficient is consistent with the omission bias account and suggests that the utility loss from losing through self-elimination compared to losing otherwise is equal to $1,056 in dollar terms.

We visually examine the model's descriptive validity in Figure 2.3. Panel A indicates that the spinning choices of Contestant 1 neatly align with the model's predictions. There

Figure 2.3: Observed and predicted spinning rates for the AQRE model with omission bias

(A) Contestant 1 **(B)** Contestant 2



Notes: The figure shows the actual spinning rate and the prediction of the AQRE model that incorporates omission bias, both for Contestant 1 (Panel A) and Contestant 2 (Panel B). All definitions are as in Figure 2.2.

is no clear underspinning or overspinning, and even the predicted spinning rate for difficult choices—first spins between 50 and 70—is close to the observed spinning rate. By contrast, however, Panel B suggests that the model underestimates Contestant 2's propensity to spin by a considerable margin, and represents Contestant 2's spinning choices distinctly less accurately than the baseline model.

The goodness-of-fit measures in Table 2.2 paint a similar picture to Figure 2.3. On the one hand, the model explains Contestant 1's decisions better than the baseline AQRE model. It correctly predicts 95.2 percent of all choices, and 83.0 percent of the difficult first spins between 50 and 70. For the baseline AQRE model, these measures are equal to 93.4 and 76.4 percent. For Contestant 2, on the other hand, the hit rate decreases from 96.5 percent in the baseline model to 95.3 percent in the present model. For difficult choices, the deterioration is even more pronounced: from 84.0 to 78.2 percent.

In the same vein, Contestant 1's overspinning rate is a mere −1.3 percentage points, which compares favorably to the −4.2 percentage points difference of the baseline model. For difficult first spins it decreases from −16.2 to −5.5 percentage points. For Contestant 2, the overspinning rate, like the hit rate, worsens: for all choices it grows from 1.9 to

4.0 percentage points, and for difficult decisions it increases from 6.8 to 14.3 percentage points.

Despite the poor descriptive accuracy of Contestant 2's choices, a Vuong test (Vuong, 1989) significantly favors the AQRE model that accounts for omission bias over the baseline AQRE model ($Z = 10.767$, $p < 0.001$).

2.6.3 Limited Foresight

Our results thus far suggest that Contestant 1 underspins compared to the model with decision errors alone, whereas Contestant 2 overspins compared to the model that combines decision errors and omission bias. In this section, we examine whether contestants' deviations from equilibrium can be explained by a fraction of decisions that is consistent with limited foresight.

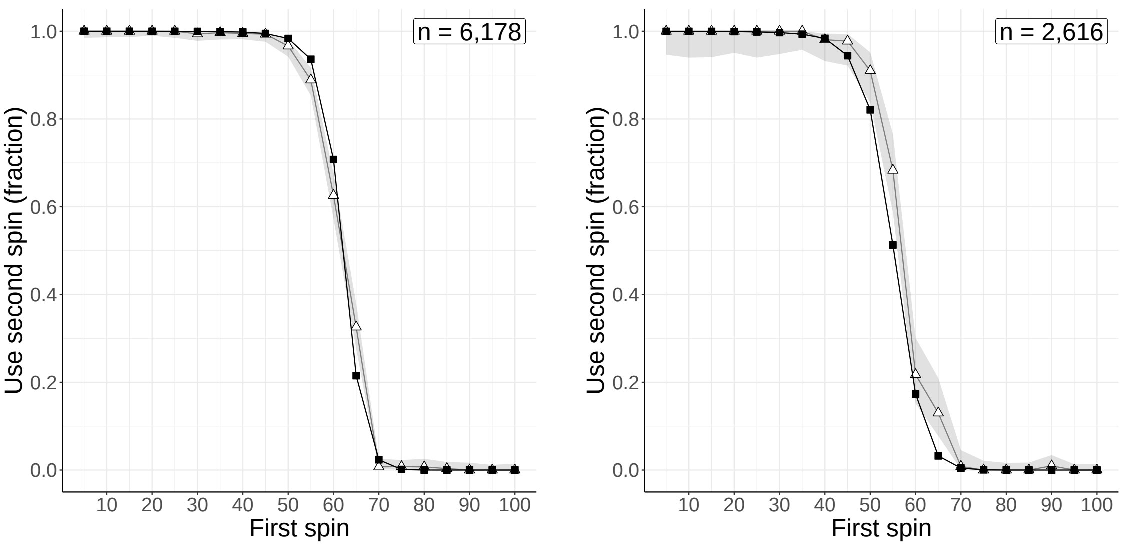
The estimation results in Table 2.2 show that the estimated value of β is 0.41, suggesting that the aggregate behavior is best explained by a fraction of 41 percent of decisions that are made in accordance with limited foresight, while the remaining 59 percent of the decisions are consistent with the correct usage of backward induction.

Figure 2.4 visually compares the predicted and observed spinning rates. Panel A suggests that the limited foresight model accurately depicts Contestant 1's spinning choices. Independent of the value of the first spin, the gap between prediction and actual observations is virtually nonexistent. Panel B shows that the model also fits the behavior of Contestant 2 reasonably well. In fact, since Contestant 2's limited foresight prediction corresponds to the backward induction prediction, Panel B of Figure 2.4 is identical to Panel B of the baseline model in Figure 2.2.

The goodness-of-fit measures in Table 2.2 similarly indicate that the predictions of the limited foresight model closely correspond to the observed behavior. The hit rate for Contestant 1's decisions is 95.2 percent (83.0 percent for difficult choices), and is identical to the AQRE model that includes omission bias. For Contestant 2, the hit rate is 96.5 percent, which is by definition equal to the baseline model and thus considerably better than the 95.3 percent hit rate of the AQRE model with omission bias. The comparison is 84.0 versus 78.2 percent for difficult choices.

Figure 2.4: Observed and predicted spinning rates for the limited foresight model

(A) Contestant 1 **(B)** Contestant 2



Notes: The figure shows the actual spinning rate and the prediction of the limited foresight model, both for Contestant 1 (Panel A) and Contestant 2 (Panel B). All definitions are as in Figure 2.2.

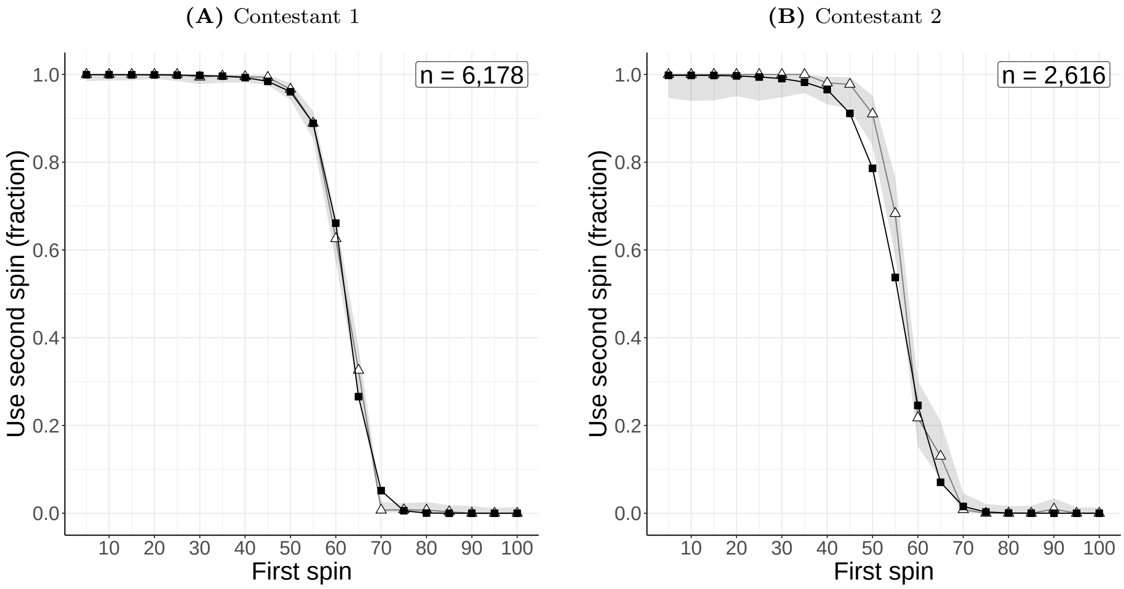
Turning to the overspinning rate, the difference between Contestant 1's average propensity to spin and the model's prediction is close to zero: 0.1 percentage points for all choices, and -0.2 percentage points for difficult first spins. For Contestant 2, the overspinning rate decreases from 4.0 percentage points in the omission bias model to 1.9 in the current, and from 14.3 to 6.8 percentage points for difficult choices.

A Vuong test confirms that the limited foresight model explains the data significantly better than both the AQRE model with omission bias ($Z = 6.157$ and $p < 0.001$) and the baseline AQRE model ($Z = 11.275$ and $p < 0.001$).

2.6.4 Omission Bias and Limited Foresight

Our last structural model incorporates omission bias in the limited foresight model. Table 2.2 shows the results. Compared to the limited foresight model without omission bias, the current model estimates a slightly higher fraction of decisions that are consistent with limited foresight: 46 versus 41 percent. Contrasting the intuition of omission bias, the estimated γ of \$173.7 is *positive*, suggesting that contestants prefer to self-eliminate over being eliminated by another contestant. Hence, any incremental improvement in the de-

Figure 2.5: Observed and predicted spinning rates for the limited foresight model with omission bias



Notes: The figure shows the actual spinning rate and the prediction of the limited foresight model that incorporates omission bias, both for Contestant 1 (Panel A) and Contestant 2 (Panel B). All definitions are as in Figure 2.2.

scriptive accuracy compared to the previous model can by definition not be explained by omission bias.

Figure 2.5 shows the observed and the predicted spinning rates. Because the predictions of the current model are almost identical to the standard limited foresight model, it also explains both contestants' choices with a reasonable degree of accuracy. At the same time, the similarity in descriptive accuracy suggests that the additional value of accounting for omission bias may be rather low.

Corroborating the visual evidence, the goodness-of-fit of the current model is highly similar to that of the standard limited foresight model. For Contestant 1, the hit rate (95.2 percent) and the overspinning rate (0.1 percentage points) are identical. The present model does provide a slight improvement in the descriptive accuracy of Contestant 2's spinning choices: the hit rate increases from 96.5 to 96.7 percent and the overspinning rate decreases from 1.9 to 1.6 percentage points. For difficult choices, the hit rate increases from 84.0 to 85.1 percent, and the overspinning rate decreases from 6.8 to 5.4 percentage points.

A Vuong test cannot reject the hypothesis that the limited foresight model with and without omission bias explain spinning choices equally well ($Z = 0.855$, $p = 0.196$).

Moreover, the positive value for the γ parameter is inconsistent with omission bias. Hence, our analysis favors the limited foresight model without omission bias.

2.7 Robustness Checks

2.7.1 Discounting the Showcase Value

Thus far, we have made two potentially important simplifications about the way in which contestants evaluate the prizes of the *Showcase* round. The first assumption is that contestants equate the utility of winning the *Showcase* round with the expected retail price of the showcase. Because the showcase consists of non-monetary prizes, a possible cause of concern is that contestants derive less utility from winning these prizes than they would from receiving the retail value in cash.

The second assumption is that contestants attach a 50 percent probability to winning the *Showcase* round if they win the SCSD. The empirical winning probability is in fact slightly lower, because neither contestant wins their showcase if both guess a value higher than the retail price.

In this section, we investigate the sensitivity of our results to the assumption that contestants discount the expected value of the showcase by 50 percent. Because contestants are assumed to be risk neutral, it is inconsequential whether this lower value results from a lower perceived showcase prize or from an equivalent reduction in the perceived probability of winning the showcase.

Table 2.4 shows the estimation results. Both the baseline AQRE model and the AQRE model that accounts for omission bias describe the observed choices less accurately than the same models in our main specification. The aggregate fit of the limited foresight models does not materially change compared to our main specification. More importantly, the ranking of the five models in terms of the descriptive accuracy remains intact. The limited foresight model provides an accurate account of contestants' spinning choices, and additionally incorporating omission bias adds little to no explanatory power (Vuong Test: $Z = 0.047$ and $p = 0.481$).

Table 2.3: Estimation results and goodness-of-fit structural models with discounted showcase value

	USPNE		Baseline		Omission bias		Limited foresight		OB & LF	
λ_1	-	-	0.0026	(0.0001)	0.0036	(0.0001)	0.0032	(0.0001)	0.0032	(0.0001)
λ_2	-	-	0.0023	(0.0002)	0.0020	(0.0001)	0.0023	(0.0001)	0.0023	(0.0001)
β	-	-	-	-	-659.4	(2.1)	-	-	-56.4	(2.1)
γ	-	-	-	-	-	-	0.479	(0.021)	0.448	(0.020)
N	8,943		8,943		8,943		8,943		8,943	
Log likelihood	-		$-1,735$		$-1,539$		$-1,452$		$-1,452$	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
Hit rate	0.934	0.967	0.934	0.965	0.944	0.953	0.952	0.965	0.952	0.961
Hit rate (50-70)	0.764	0.851	0.764	0.840	0.798	0.782	0.830	0.840	0.830	0.823
Overspinning	-0.063	-0.002	-0.055	0.017	-0.036	0.030	0.002	0.017	-0.000	0.018
Overspinning (50-70)	-0.234	-0.004	-0.204	0.084	-0.132	0.144	0.006	0.084	-0.000	0.089

Notes: The table shows the results and goodness-of-fit for various models of strategic decision making. Contestants are assumed to discount the monetary value of the showcase by 50 percent. All definitions are as in Table 2.2.

2.7.2 Ignoring Bonus Prizes

It is conceivable that contestants simplify their spinning choice by ignoring the possibility of winning bonus prizes by scoring exactly 100 points, and instead only focus on winning the SCSD. Doing so would effectively reduce their incentive to spin. In this section, we estimate the structural models under the assumption that contestants ignore the possibility of winning bonuses. Because the remaining outcome variable—winning or not winning the SCSD—is binary, this analysis has the additional benefit that risk preferences by definition do not play a role.

In line with our main findings, the estimation results in Section 2.7.2 show that neither decision errors nor omission bias can adequately explain the observed spinning choices, and that limited foresight provides the most accurate description of contestants' behavior. There is no significant improvement in the limited foresight model's ability to capture contestants' decisions when we additionally account for omission bias (Vuong test: $Z = -0.767$ and $p = 0.778$).

2.8 Learning

Because the show has been running uninterruptedly since 1972, an interesting question is whether the fraction of contestants that follow the optimal strategy increases over time. To examine the effect of learning, we estimate the limited foresight model for four different time periods: (i) from 1979 to 2001, (ii) from 2002 to mid-2008, (iii) from mid-2008 to

Table 2.4: Estimation results and goodness-of-fit structural models in which contestants ignore the bonus prizes

	USPNE		Baseline		Omission bias		Limited foresight		OB & LF	
λ_1	-	-	0.0015	(0.0000)	0.0019	(0.0001)	0.0017	(0.0001)	0.0016	(0.0001)
λ_2	-	-	0.0011	(0.0000)	0.001	(0.0000)	0.0011	(0.0000)	0.0011	(0.0000)
γ	-	-	-	-	−794.8238	(2.2)	-	-	173.6714	(2.9)
β	-	-	-	-	-	-	0.3407	(0.0190)	0.4571	(0.031)
N	8,943		8,943		8,943		8,943		8,943	
Log likelihood	-		−1,635		−1,557		−1,467		−1,453	
	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
Hit rate	0.934	0.953	0.934	0.953	0.952	0.953	0.952	0.953	0.952	0.953
Hit rate (50-70)	0.764	0.782	0.764	0.782	0.830	0.782	0.830	0.782	0.830	0.784
Overspinning	−0.063	0.037	−0.036	0.025	−0.013	0.041	0.001	0.025	0.010	0.022
Overspinning (50-70)	−0.234	0.184	−0.142	0.091	−0.057	0.145	−0.001	0.092	0.029	0.078

Notes: The table shows the results and goodness-of-fit for various models of strategic decision making. Contestants are assumed to ignore the bonus prizes that can be obtained from spinning exactly 100 points. All definitions are as in Table 2.2.

Table 2.5: Estimation results and goodness-of-fit limited foresight model for different periods

	1979 to 2001		2002 to mid-2008		mid-2008 to 2013		2014-2018	
λ_1	0.0014	(0.0001)	0.0017	(0.0001)	0.0018	(0.0001)	0.0018	(0.0001)
λ_2	0.0011	(0.0001)	9e-04	(0.0001)	0.0013	(0.0001)	0.0013	(0.0001)
β	0.455	(0.042)	0.434	(0.038)	0.411	(0.036)	0.320	(0.042)
N	2,421		2,080		2,561		1,881	
Log likelihood	−431		−302		−390		−317	
	C1	C2	C1	C2	C1	C2	C1	C2
Hit rate	0.944	0.970	0.959	0.956	0.954	0.964	0.952	0.968
Hit rate (50-70)	0.810	0.884	0.846	0.786	0.833	0.834	0.833	0.841
Overspinning	0.001	0.016	0.000	0.017	-0.000	0.013	0.001	0.032
Overspinning (50-70)	-0.007	0.049	-0.001	0.064	-0.001	0.040	-0.004	0.144

Notes: The table shows the results for the limited foresight model for four different time periods. All definitions are as in Table 2.2.

2013, and (iv) from 2014 to 2018. The results in Table 2.5 provide suggestive evidence that learning takes place, albeit to a limited degree. Between 1979 and 2001, the model estimates that 55 percent of the choices are consistent with backward induction. Between 2014 to 2018, this fraction has increased to 68 percent. The rationality parameter λ_1 correspondingly increases over time. Yet, despite the effect of learning and the accessibility of the optimal strategy, deviations from equilibrium remain persistent.

2.9 Conclusion and Discussion

The present chapter examines the optimality of strategic decisions in the *Showcase Showdown* (SCSD), a simple sequential game of perfect information in the long-running American TV show *The Price is Right.* The unique subgame perfect Nash equilibrium (USPNE) for this game can be found through backward induction. The high stakes, relatively simple choice environment and ample learning opportunities provide a particularly benign setting for the game-theoretic predictions to hold.

Despite the conducive setting, however, our analysis shows that contestants' spinning decisions often deviate from the equilibrium strategy. These deviations cannot be explained by random decision errors, or a combination of decision errors and a preference for harm caused by a failure to act over the same harm caused by an explicit action. Instead, the observed behavior can be explained by limited foresight, where a contestant myopically bases her decision only on the next stage of the game. Although more and more contestants appear to follow the optimal strategy over time, deviations remain commonplace even after several decades of *The Price is Right*.

At first sight, the prevalence of these deviations may seem remarkable given the substantial gains that contestants forego in expectation while the optimal strategy is publicly available. Such ignorance may in fact be rational, however, if the search costs of discovering the optimal strategy outweigh the expected benefits (Stigler, 1961). The *ex ante* odds that a given studio audience member actually gains from knowing the optimal strategy are in fact rather slim. To participate in a *Price is Right* episode, audience members first need to be selected from a studio audience with 324 other people. Once selected, they need to win an auction game against four others to proceed to the SCSD. In this round, knowing the optimal strategy only contributes to her winning chance in the relatively rare situations where her 'naive' strategy would be different from the optimal one: looking up the optimal strategy only pays off if she wins the SCSD *because* she modified her strategy. As a consequence, the expected benefits of knowing the optimal strategy are minimal for any given audience member. Hence, despite the valuable prizes that can be won, refraining from searching for the optimal strategy before attending the show can easily be justified by the negligible increase of the winning chance it yields.

Independent of the nature of the ignorance—rational or otherwise—contestants seem to resort to simpler decision rules than backward induction to devise their spinning strategies. The fact that this type of behavior arises in the SCSD is at odds with experimental research that shows how failures of backward induction tend to disappear when players play relatively simple games (Dufwenberg et al., 2010; Ho and Su, 2013; Dufwenberg and Van Essen, 2018), face higher stakes (Rapoport et al., 2003) or get sufficient opportunity to learn the optimal strategy (Dufwenberg et al., 2010; Gneezy et al., 2010). Our results indicate that even the fulfillment of all three conditions may still be insufficient to ensure the descriptive validity of the game-theoretic predictions.

A potential external validity threat is that contestants' decisions are anomalous because they are being watched by millions of viewers at home and in the studio. In theory, being in the limelight may distract or fluster contestants, and induce a relatively high number of mistakes compared to more anonymous settings. Alternatively, contestants may have a preference for being on screen for as long (or short) as possible. We cannot fully dismiss the relevance of these considerations, but there are at least three reasons why our findings are unlikely to be driven by public scrutiny. First, Tenorio and Cason (2002) show that there is no clear difference between spinning choices in the TV show and decisions in the laboratory. Second, Baltussen et al. (2016) show that the general pattern of behavior in the limelight of a simulated game show environment is relatively similar to that in an anonymous computerized laboratory setting. Third, our analysis of omission bias implicitly also takes into account screen time preferences, because the spinning choices that prevent self-elimination are identical to those that prolong screen time. Because the general pattern of behavior cannot be explained by omission bias, the same conclusion applies to screen time preferences.

A more serious concern is that attendees of *The Price is Right* are unrepresentative for the general population. Prior studies show that people who are highly strategically sophisticated—chess and GO players, as well as US senators—are relatively likely to follow the game-theoretic predictions (Palacios-Huerta and Volij, 2009; Baghestanian and Frey, 2016; Spenkuch et al., 2018). Yet, because *The Price is Right* contestants are comparatively diverse in terms of age, gender, ethnicity and education levels, they form a more accurate cross-section of the population than senators or chess and GO players. Hence, while highly strategically sophisticated people may behave in accordance to backward induction, our results suggest that the general population may not.

2.A Appendix

Table 2.6: Expected loss for deviations from the USPNE in 2018

First spin	Contestant 1		Contestant 2	
5	$2,905.09	(Spin)	$4,643.03	(Spin)
10	$2,890.26	(Spin)	$4,554.06	(Spin)
15	$2,862.11	(Spin)	$4,391.06	(Spin)
20	$2,818.06	(Spin)	$4,150.7	(Spin)
25	$2,754.41	(Spin)	$3,829.68	(Spin)
30	$2,666.25	(Spin)	$3,424.64	(Spin)
35	$2,547.46	(Spin)	$2,932.28	(Spin)
40	$2,390.58	(Spin)	$2,349.26	(Spin)
45	$2,186.87	(Spin)	$1,672.25	(Spin)
50	$1,926.13	(Spin)	$897.94	(Spin)
55	$1,606.48	(Spin)	$57.91	(Spin)
60	$1,064.27	(Spin)	$884.41	(Not spin)
65	$358.22	(Spin)	$1,932.36	(Not spin)
70	$483.14	(Not spin)	$3,089.27	(Not spin)
75	$1,584.12	(Not spin)	$4,358.46	(Not spin)
80	$2,947.16	(Not spin)	$5,743.27	(Not spin)
85	$4,613.49	(Not spin)	$7,247	(Not spin)
90	$6,627.89	(Not spin)	$8,873	(Not spin)
95	$9,038.83	(Not spin)	$10,624.59	(Not spin)
100	$11,898.53	(Not spin)	$12,658.48	(Not spin)

Notes: The table displays the expected loss for deviations from the USPNE in 2018 if all other contestants follow the optimal strategy. *First spin* is the value of the first spin. *Contestant 1* is the expected loss when Contestant 1 deviates from the USPNE. *Contestant 2* is the expected loss when Contestant 2's deviates from the USPNE after beating the score of Contestant 1 with her first spin. The optimal strategy is given in parentheses.

Table 2.7: USPNE per year

Year	E_s	Bonus scheme	Contestant 1	Contestant 2
1979	$8,369	1	65	55
1980	$9,981	1	65	55
1981	$10,903	1	65	55
1982	$10,556	1	65	55
1983	$10,929	1	65	55
1984	$11,446	1	65	55
1985	$11,735	1	65	55
1986	$13,087	1	65	55
1987	$14,523	1	65	55
1988	$16,617	1	65	55
1989	$18,063	1	65	55
1990	$19,059	1	65	55
1991	$18,779	1	65	55
1992	$19,677	1	65	55
1993	$20,273	1	65	55
1994	$21,246	1	65	55
1995	$21,853	1	65	55
1996	$22,908	1	65	55
1997	$23,140	1	65	55
1998	$23,447	1	65	55
1999	$23,606	1	65	55
2000	$24,879	1	65	55
2001	$25,995	1	65	55
2002	$27,882	1	65	50
2003	$31,021	1	65	50
2004	$34,482	1	65	50
2005	$34,287	1	65	50
2006	$33,326	1	65	50
2007	$30,845	1	65	50
2008 (until September 21)	$30,198	1	65	50
2008 (from September 22)	$26,502	2	65	55
2009	$27,524	2	65	55
2010	$28,258	2	65	55
2011	$29,286	2	65	55
2012	$30,029	2	65	55
2013	$29,788	2	65	55
2014	$30,147	2	65	55
2015	$29,655	2	65	55
2016	$28,612	2	65	55
2017	$28,660	2	65	55
2018	$28,965	2	65	55

Notes: The table shows the optimal spinning threshold per year for Contestant 1 and 2. E_s is the expected showcase value. *Bonus scheme* shows the prevailing bonus scheme. The prizes in Bonus Scheme 1 are equal to $1,000, $5,000 and $10,000, and the prizes in Bonus Scheme 2 are equal to $1,000, $10,000 and $25,000 respectively. *Contestant 1* shows the maximum score with which Contestant 1 should spin again, and *Contestant 2* shows the analogous threshold for Contestant 2 if her first spin beats the score of Contestant 1.

Chapter 3

Does Losing Lead to Winning?

An Empirical Analysis for Four Different Sports[1]

3.1 Introduction

In an influential paper, Berger and Pope (2011, henceforth BP) argue that lagging behind halfway through a competition does not necessarily imply a lower likelihood of winning, and that being slightly behind can actually increase the chance to come out on top. In particular, they argue that because winning is the goal, the performance of an opponent will serve as a salient benchmark—or reference point—to which a competitor will compare his or her own performance during the competition. Research on goals as reference points shows that people who are slightly below their goal work harder than those who have reached or exceeded it (Heath et al., 1999; Pope and Simonsohn, 2011; Corgnet et al., 2015; Allen et al., 2016). Analogously, BP argue that people who are slightly behind in a competition may be more motivated than people who are slightly ahead.

To test this hypothesis, BP analyze more than sixty thousand professional and collegiate basketball matches. Their main analyses focus on the score difference at half-time, because the relatively long break that occurs halfway through the match provides players with the time to reflect on their position relative to their opponent. BP find that National Basketball Association (NBA) teams that are slightly behind are between 5.8 and 8.0 percentage points more likely to win the match than those that are slightly ahead.

[1] This chapter is based on a working paper titled *Does Losing Lead to Winning? An Empirical Analysis for Four Different Sports* and is joint work with Martijn van den Assem and Dennie van Dolder.

For collegiate matches, they similarly find a positive effect of being behind, but the size of the effect is smaller.

The present chapter extends the analysis of BP to large samples of Australian football, American football and rugby matches, and revisits the analysis of basketball. Our main analyses consider the effect of being slightly behind at half-time on the likelihood of winning the match. To estimate this effect, we use a regression discontinuity design (RDD; Thistlethwaite and Campbell, 1960). Whenever possible, we in addition analyze whether marginally trailing at half-time improves the likelihood of winning the third or fourth quarter separately, as was also done by BP, and whether marginally trailing after the third quarter improves the odds of winning the match.

For Australian football, American football and rugby we find little to no evidence that being slightly behind improves performance. For basketball we replicate the finding that half-time trailing in NBA matches from the period analyzed in BP improves the odds of winning. Our estimated effect size of 8.3 percentage points is even somewhat larger than the effect size reported in BP. For NBA matches from outside that period, for collegiate matches, and for matches from the Women's National Basketball Association (WNBA), however, we obtain null results.

To synthesize our results, we conduct a meta-analysis that estimates the overall effect of trailing at half-time on winning the match across all competitions and sports. The estimated meta-analytic effect is close to zero and statistically insignificant. The narrow confidence interval for the overall effect size suggests that the true effect, if existent at all, is likely relatively small. Similar conclusions follow from meta-analyses of the effect of half-time trailing on winning the third or fourth quarter separately, and from a meta-analysis of the effect of trailing after the third quarter on winning the match.

The chapter proceeds as follows. Section 3.2 explains the methodology, Sections 3.3 to 3.6 show the results for each of the four sports, Section 3.7 presents the meta-analyses, and Section 3.8 discusses our findings and concludes.

3.2 Empirical Strategy

We employ a regression discontinuity design (RDD) to estimate the causal impact of being behind on performance. RDDs are used to estimate treatment effects in non-experimental settings. The distinct feature is that the treatment is assigned based on whether an observed covariate, the so-called running variable, exceeds a specific cutoff value. Under the testable assumption that all other determinants of the outcome variable are continuous through this cutoff value, the variation in the treatment status is "as good as randomized by an experiment" (Lee, 2008, p.676), and a discontinuity in the outcome variable at the cutoff can be causally attributed to the treatment.

In our main analyses, the running variable is the score difference at half-time and the cutoff value is zero. We estimate the following regression model:

$$Y_i = \alpha + \tau \times T_i + \beta_1 \times X_i + \beta_2 \times T_i \times X_i + \varepsilon_i \tag{3.1}$$

where Y_i is an indicator variable that takes the value of 1 if team i wins the match, and X_i is the half-time score difference between team i and the opposing team. The treatment variable T_i takes the value of 1 if team i is behind at half-time. The coefficient τ represents the discontinuity in the winning probability at a zero score difference. Under the hypothesis that being slightly behind improves performance, this coefficient is positive. The interaction term $T_i \times X_i$ allows for different slopes above and below the cutoff. To not use every match twice, we systematically take the perspective of the home team. We omit matches where teams were tied at half-time.[2]

If the assumption of a piece-wise linear relation between the winning probability and the half-time score difference is violated, the regression model will generate a biased estimate of the treatment effect. As a solution to this problem, Hahn et al. (2001) propose the use of local-linear regression. Even if the true relation is non-linear, a linear specification can provide a close approximation within a limited bandwidth around the cutoff.

[2] If the score difference is negative, the home team is treated and the away time is not, whereas if the score difference is positive, the away team is treated and the home team is not. In matches with a zero half-time score difference, in contrast, neither of the teams is treated. These matches can therefore neither be used to estimate the linear relation below the cutoff value, nor to estimate the linear relation above it.

A downside of this solution is that it reduces the effective number of observations, and therefore the precision of the estimate. To strike the appropriate balance between bias and precision, we use the local-linear method proposed by Calonico et al. (2014). Their method selects the bandwidth that minimizes the mean squared error, corrects the estimated treatment effect for any remaining non-linearities within the bandwidth, and linearly downweights observations that are farther away from the cutoff.

Our RDD requires that the skill difference between home and away teams is continuous through the cutoff. To examine whether this assumption holds, we in addition estimate a modified version of Equation (3.1), where the outcome variable is the skill difference between the two teams. As a proxy for the skill difference, we use the difference between the proportions of home matches won by the home team and away matches won by the away team during the current calendar year.[3] We again employ the local-linear method proposed by Calonico et al. (2014).

We consider data from four different sports: Australian football, American football, rugby and basketball. In all these sports, teams generally score a large number of points. The validity of our RDD hinges on the assumption of a piece-wise linear relation between the full-time winning probability and the half-time score difference within a bandwidth around the cutoff. In sports where teams typically score only a small number of points, even the smallest possible half-time disadvantage has a strong impact on the probability of losing the match, and the marginal effect of larger differences quickly converges to zero. Consequently, for low-scoring sports, the assumption of linearity is violated even within a small bandwidth around the cutoff. Also, and perhaps even more importantly, the hypothesized psychological effect is unlikely to occur in such sports: when being behind is relatively hard to overcome, trailing by one or a few points likely discourages rather than motivates (Fershtman and Gneezy, 2011; Gill and Prowse, 2012).

Matches have to satisfy a number of criteria for inclusion. First, the half-time score, the full-time score and the year of play need to be available. Second, the match must not have ended in a draw. Last, the match must not be the only home (away) match played by the home (away) team in the given year. The latter condition is necessary for our test

[3] The current match itself is excluded from these calculations. We exclusively use home (away) matches for the home (away) team to take account of the home advantage and possible imbalances in the numbers of home and away matches.

of the assumption that the skill difference between home and away teams is continuous at the cutoff.

We present our results on a sport-by-sport basis. For each sport, we look at multiple competitions. We always start with graphs that show the proportion of matches won by home teams at given half-time score differences. We construct these graphs following the approach proposed by Calonico et al. (2015). In each graph, smooth curves on both sides of the cutoff give a visual impression of whether the relation is approximately linear within the estimated bandwidth, and provide a first indication of the existence of a discontinuity. Next, we present the results for the main RDD, where the outcome variable takes the value of 1 if the home team won the match, and where the running variable is the score difference at half-time. To assess the robustness of the results, we examine the sensitivity of the estimated coefficients to a range of imposed alternative bandwidths. If matches of a sport consist of quarters and we have data on the score after the third quarter, we in addition analyze the effect of trailing at half-time on winning the third quarter and on winning the fourth quarter separately, as well as the effect of trailing after the third quarter on winning the match. Last, for each RDD, we examine the assumption that the skill difference is continuous through the cutoff.

3.3 Australian Football

3.3.1 Description and Data

The first sport that we consider is Australian football. We use data from two different leagues. One is the Australian football League (AFL), which is widely considered to be the sport's most important league. It is the only fully professional Australian football division and the fourth most popular sports competition in the world by average weekly attendance.[4] The other is the South Australian National Football League (SANFL). The SANFL is a semi-professional regional football league played in South Australia.

[4]The Guardian. 2014. Battle of the codes: Australia's four sports leagues compared. Available from https://www.theguardian.com/news/datablog/interactive/2014/apr/15/australia-football-interactive-statistics [Accessed: 7 July 2020].

Table 3.1: Summary statistics Australian football

Panel A: AFL (1897-2018, N=14,945)

	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	84.4	25.3	10	67	84	101	210
Total points at full-time	171.1	46.3	24	140	172	202	345
Score difference at half-time	4.4	24.2	−107	−11	5	20	120
Score difference at full-time	8.9	40.7	−164	−18	9	35	190
Home team wins match	0.60	0.49	0	0	1	1	1

Panel B: SANFL (1950-2018, N=6,622)

	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	90.1	24.9	20	73	89	106	216
Total points at full-time	184.6	44.8	38	153	183	214	396
Score difference at half-time	4.3	27.4	−112	−14	5	22	108
Score difference at full-time	8.9	48.0	−178	−23	10	40	238
Home team wins match	0.58	0.49	0	0	1	1	1

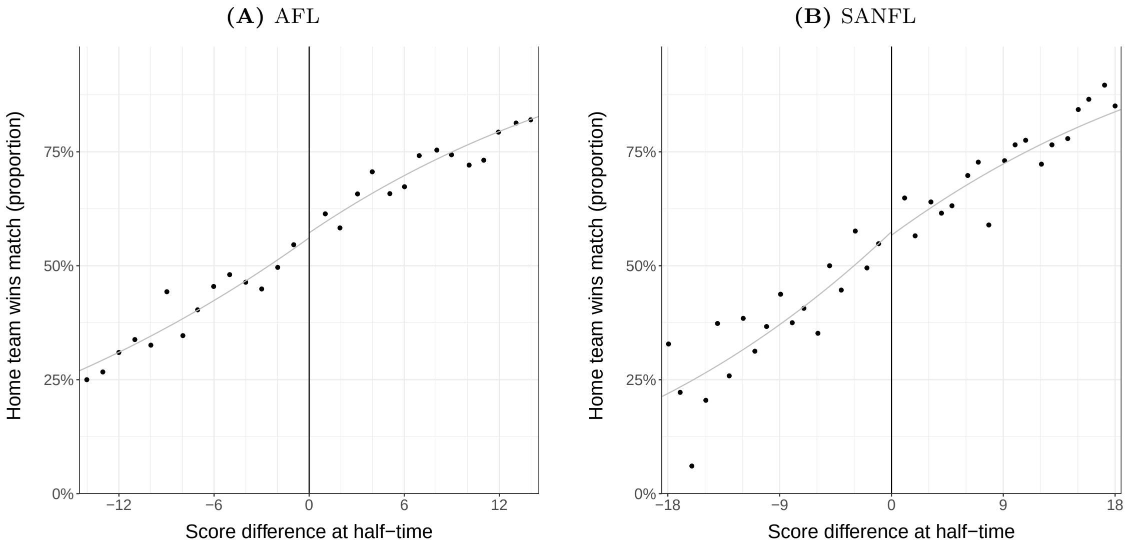
Notes: The table displays the summary statistics for AFL and SANFL matches where the half-time score difference was not zero. *Home team wins match* is an indicator variable that takes the value of 1 if the home team won the match. *Total points at half-time (full-time)* is the total number of points scored by the two teams together at half-time (full-time). *Score difference at half-time (full-time)* is the half-time (full-time) score difference between the home and away team.

Australian football is played by two teams of 18 players each on an oval shaped pitch. At both ends of the field there are four goal posts behind a goal line. The object of the game is to kick the ball between the posts. A team scores six points by kicking the ball between the middle two posts. Teams score one point (i) when they kick the ball between a middle post and one of the outside posts, (ii) when a player on the ground touched the ball before it goes between the middle posts, and (iii) when a defender is forced to carry the ball across its own goal line. Australian football matches consist of four 20-minute quarters. There is a 20-minute half-time break and there are two 6-minute breaks after the first and third quarter.

We obtained data for 15,209 AFL and 6,728 SANFL matches that satisfy the criteria stipulated in Section 3.2.[5] The exclusion of matches with a zero half-time score difference reduces these samples to 14,945 (AFL) and 6,622 (SANFL) matches. Table 3.1 summarizes the data. On average, the two teams together scored 171 points in AFL matches and 185 in SANFL matches. At half-time, these numbers were 84 and 90, respectively.

[5]We scraped the AFL matches from www.afltables.com on 3 September 2018 and the SANFL matches from www.australianfootball.com on 2 October 2018. The data on both websites are collected and edited by fans.

Figure 3.1: Regression discontinuity plots for Australian football



Notes: The figure shows the regression discontinuity plots for AFL (Panel A) and SANFL (Panel B) matches with a half-time score difference that was within a limited bandwidth around the cutoff value of zero. The plots are constructed using the approach proposed by Calonico et al. (2015). Each dot represents the proportion of matches won by the home team at a given half-time score difference. The curves on both sides of the cutoff are fourth-order polynomials. The bandwidths correspond to the bandwidth estimates deriving from our main regression discontinuity design.

In both samples, home teams on average led by 4 points at half-time and by 9 points at full-time, and won roughly 60 percent of the times.

3.3.2 Analysis and Results

We first visually explore the relation between the half-time score difference and the full-time winning probability. Figure 3.1 shows that the relation is approximately linear on both sides of the cutoff value of zero, both for AFL and for SANFL matches. The winning probability increases at a rate of roughly two percentage points per point. There is no clear evidence of a discontinuous change at the cutoff.

Table 3.2, Panel A presents the results for the main RDD. There is no evidence of a positive performance effect from trailing. In fact, the point estimate for AFL teams indicates that being slightly behind at half-time *decreases* the winning chances by 3.4 percentage points, but this effect is statistically insignificant ($p = 0.253$). For the SANFL sample, the point estimate of the effect of being behind is virtually zero ($p = 1.000$). The wide 95 percent confidence intervals for the two estimates, however, indicate that a

Table 3.2: Results for Australian football

	AFL	SANFL
Panel A: Score difference at half-time, winning match		
Behind at half-time	−0.034	0.000
	(−0.092, 0.024)	(−0.082, 0.082)
Bandwidth	14.73	18.53
Total observations	14,945	6,622
Included observations	6,902	3,348
Panel B: Score difference at half-time, winning third quarter		
Behind at half-time	−0.004	0.027
	(−0.053, 0.046)	(−0.054, 0.108)
Bandwidth	24.23	23.66
Total observations	14,599	6,471
Included observations	10,124	3,990
Panel C: Score difference at half-time, winning fourth quarter		
Behind at half-time	−0.028	0.004
	(−0.084, 0.027)	(−0.078, 0.087)
Bandwidth	19.61	23.04
Total observations	14,615	6,491
Included observations	8,577	3,997
Panel D: Score difference after third quarter, winning match		
Behind after third quarter	0.016	−0.020
	(−0.059, 0.091)	(−0.122, 0.083)
Bandwidth	12.04	16.55
Total observations	15,040	6,655
Included observations	4,447	2,230

Notes: The table reports the estimated effect of being behind on the likelihood of winning for AFL and SANFL matches using a regression discontinuity design. Treatment effects are estimated with the local-linear non-parametric estimator proposed by Calonico et al. (2014). The outcome variable is *Home team wins match* (Panels A and D), *Home team wins third quarter* (Panel B) or *Home team wins fourth quarter* (Panel C). The running variable is *Score difference at half-time* (Panels A, B and C) or *Score difference after third quarter* (Panel D). *Bandwidth* is the largest absolute score difference for matches included in the RDD. *Total observations* is the number of observations in the analyzed sample. *Included observations* is the number of observations within the bandwidth. Numbers in parentheses represent 95 percent confidence intervals. Asterisks denote significance at the 0.01 (***), 0.05 (**) and 0.1 (*) level.

considerable range of positive and negative effect sizes cannot be ruled out. Figure 3.6 in the Appendix shows that the results are robust to a range of imposed alternative bandwidths.

A possible explanation for the absence of evidence of a performance-enhancing effect could be that the effect is too ephemeral to materially affect the full-time match outcome. If being behind at half-time improves performance only temporarily, we are more likely to find an effect in a shorter period directly following the half-time break. We therefore also analyze the effect of being behind on performance in the third quarter separately. For completeness, we also look at the effect on the fourth quarter. In these alternative RDDs, the outcome variable takes the value of 1 if the home team scored more points than the away team in the given quarter. We again exclude matches where the half-time score difference was zero, and now also omit matches where both teams scored the same number of points in the quarter of interest. Panels B and C show the results. Mirroring the picture emerging from the main RDD, there is no statistically significant evidence that trailing at half-time affects performance in the next (third) quarter (AFL: $\tau = -0.004, p = 0.889$; SANFL: $\tau = 0.027, p = 0.508$). Not surprisingly, the estimated treatment effect in the final (fourth) quarter is also insignificant (AFL: $\tau = -0.028, p = 0.318$; SANFL: $\tau = 0.004, p = 0.918$).

Being slightly behind is potentially more consequential in later stages of the match. We therefore also analyze whether being behind after the third quarter improves performance in the final quarter. In these alternative RDDs, winning the match is the outcome variable and the score difference after the third quarter is the running variable. For this analysis we include matches with a zero score difference at half-time and exclude matches with a zero score difference after the third quarter. Panel D shows the results. The estimates for the AFL ($\tau = 0.016, p = 0.674$) and SANFL ($\tau = -0.020, p = 0.707$) samples are both insignificant.

Last, to investigate the validity of the RDDs, we examine the identifying assumption that the skill difference between the home and away team is continuous at the cutoff value of a zero score difference. There is no significant evidence for a discontinuity, neither at half-time (AFL: $p = 0.217$; SANFL: $p = 0.633$) nor at the end of the third quarter (AFL: $p = 0.650$; SANFL: $p = 0.177$).

Taken together, the results for Australian football do not support the hypothesis that being slightly behind increases the odds of winning. We cannot reject the null hypothesis of no effect, neither in the two main analyses nor in the additional analyses.

3.4 American Football

3.4.1 Description and Data

The second sport that we consider is American football. We analyze matches from the National Football League (NFL) and from the National Collegiate Athletic Association (NCAA). The NFL is seen as the most important American football league and is the best attended professional sports league in the world.[6] The NCAA is an American inter-collegiate competition, and is often regarded as the second-highest level of competition in American football.

American football matches are played between two teams of 11 players. The playing field is rectangular and contains an end zone on each side. In each end zone, there are two posts with a crossbar. Teams score a touchdown, worth six points, when a player either catches the ball within the opposing team's end zone or advances into the end zone while holding the ball. After a touchdown, the offensive team gets the opportunity to score either one point by kicking the ball through the posts from a distance of 15 yards from the end zone, or two points by taking the ball into the end zone from a distance of two (NFL) or three (NCAA) yards from the end zone. A team scores a field goal, worth three points, by kicking the ball through the posts during normal play. The defensive team gains two points when they tackle a member of the opposing team who holds the ball in the opposing team's end zone. Matches consist of four 15-minute quarters. There is a 12-minute half-time break and there are two 2-minute breaks after the first and third quarter.

We obtained data for 12,279 NFL and 7,936 NCAA matches that satisfy the data requirements outlined in Section 3.2.[7] The exclusion of matches with a zero half-time score difference reduces the samples to 10,590 (NFL) and 7,024 (NCAA) matches. Table 3.3 shows summary statistics. Together, teams scored on average 41 (NFL) and 55 (NCAA) points per match. At half-time, these numbers were 21 and 29, respectively. Home teams

[6]Business Insider. 2015. The NFL and Major League Baseball are the most attended sports leagues in the world. Available from: https://www.businessinsider.com/attendance-sports-leagues-world-2015-5 [Accessed: 7 July 2020].

[7]We scraped the NFL data from www.pro-football-reference.com on 8 September 2018, and the NCAA data from www.sports-reference.com on 2 October 2018. Both websites report official NFL and NCAA statistics.

Table 3.3: Summary statistics American football

	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Panel A: NFL (1945-2017, N=10,590)							
Total points at half-time	20.7	8.7	2	14	20	27	62
Total points at full-time	40.6	12.1	8	32	41	49	113
Score difference at half-time	2.0	11.5	−35	−7	3	10	42
Score difference at full-time	2.9	15.5	−55	−7	3	14	59
Home team wins match	0.58	0.49	0	0	1	1	1
Panel B: NCAA (2003-2018, N=7,024)							
	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	29.0	11.8	2	21	28	37	94
Total points at full-time	55.0	17.6	5	43	54	66	137
Score difference at half-time	4.3	15.2	−49	−7	5	14	56
Score difference at full-time	7.1	21.5	−73	−7	7	22	78
Home team wins match	0.63	0.48	0	0	1	1	1

Notes: The table displays the summary statistics for NFL and NCAA matches where the half-time score difference was not zero. All definitions are as in Table 3.1.

led by an average of 2 (NFL) and 4 (NCAA) points at half-time, and 3 (NFL) and 7 (NCAA) points at full-time. Home teams won roughly 60 percent of the matches.

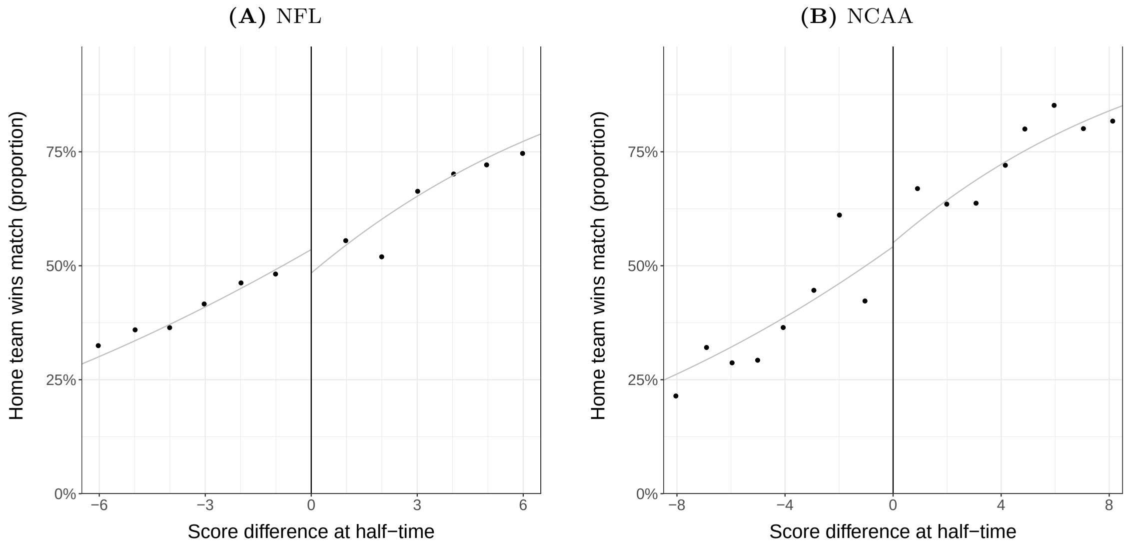
3.4.2 Analysis and Results

Figure 3.2 shows that the proportion of home teams winning the match increases approximately linearly with the half-time score difference, at a rate of roughly four percentage points per point for both NFL and NCAA matches. There appears to be a small negative discontinuity at zero for NFL teams, suggesting that marginally trailing at half-time enhances teams' performance. In NCAA matches, there is no indication of such an effect.

Table 3.4, Panel A presents the results for the main RDD. Statistically there is no significant evidence that half-time trailing affects the full-time winning probability, neither for NFL ($p = 0.431$) nor for NCAA ($p = 0.425$) matches. The estimated effect sizes of 4.8 (NFL) and −4.6 (NCAA) percentage points are economically substantial, but the 95 percent confidence intervals are wide. Figure 3.7 in the Appendix shows that the results are robust to alternative bandwidth choices.

The effect of trailing at half-time may be relatively short-lived. We therefore also analyze the effect of half-time trailing on performance in the third (and fourth) quarter separately. The outcome variable now takes the value of 1 if the team scored more points than the opposing team in the quarter of interest. In addition to excluding matches

Figure 3.2: Regression discontinuity plots for American football



Notes: The figure shows the regression discontinuity plots for NFL (Panel A) and NCAA (Panel B) matches. Definitions are as in Figure 3.1.

where the half-time score difference was zero, we now also exclude matches where the two teams scored the same number of points in the quarter of interest. Panels B and C show the results. In contrast to the results for the main RDD, being slightly behind at half-time in the NFL does significantly improve performance in the first quarter after the break: half-time trailing increases the odds of winning the third quarter by a sizable 13.4 percentage points ($p = 0.018$). There is no statistically significant evidence for such an effect in the third quarter for NCAA matches, but the point estimate is economically large ($\tau = 0.072, p = 0.206$). The effect of half-time trailing on the odds of winning the fourth quarter is insignificant in both samples (NFL: $\tau = -0.000, p = 0.997$; NCAA: $\tau = -0.106, p = 0.158$).

To further examine whether the effect exists within a single quarter, we examine the effect of trailing after the third quarter on the likelihood of winning the match. We now include matches with a zero score difference at half-time, and exclude matches with a zero score difference after the third quarter.

Panel D shows the results. The two estimates are statistically insignificant (NFL: $\tau = -0.052, p = 0.507$; NCAA: $\tau = -0.018, p = 0.855$), which suggests that trailing after the third quarter does not materially affect the chance of winning the match.

Table 3.4: Results for American football

	NFL	NCAA
Panel A: Score difference at half-time, winning match		
Behind at half-time	0.048	−0.046
	(−0.072, 0.168)	(−0.160, 0.067)
Bandwidth	6.06	8.70
Total observations	10,590	7,024
Included observations	3,736	2,812
Panel B: Score difference at half-time, winning third quarter		
Behind at half-time	0.134**	0.072
	(0.023, 0.245)	(−0.040, 0.185)
Bandwidth	6.27	8.83
Total observations	10,107	6,712
Included observations	3,557	2,592
Panel C: Score difference at half-time, winning fourth		
Behind at half-time	-0.000	−0.106
	(−0.125, 0.124)	(−0.252, 0.041)
Bandwidth	5.59	7.00
Total observations	10,187	6,827
Included observations	3,040	1,695
Panel D: Score difference after third quarter, winning match		
Behind after Q3	−0.052	−0.018
	(−0.207, 0.102)	(−0.208, 0.173)
Bandwidth	6.31	5.41
Total observations	8,956	6,235
Included observations	3,678	1,845

Notes: The table reports the estimated effect of being behind on the likelihood of winning for NFL and NCAA matches using a regression discontinuity design. Definitions are as in Table 3.2.

Last, we examine whether the identifying assumption that the skill difference between home and away teams is continuous through the cutoff holds. There is no significant evidence of a discontinuity, neither at half-time (NFL: $p = 0.604$; NCAA: $p = 0.218$) nor after the third quarter (NFL: $p = 0.311$; NCAA: $p = 0.649$).

Overall, our analyses of American football provide little evidence that being behind improves performance. The only exception is that being behind at half-time in the NFL has a significantly positive effect on the chances of winning the third quarter. The other analyses for the NFL and the analyses for the NCAA do not provide supportive evidence for the hypothesis that trailing enhances performance.

3.5 Rugby

3.5.1 Description and Data

The third sport that we analyze is rugby. There are two similar yet distinct forms, namely rugby union and rugby league. For rugby union, our analysis covers international matches, including matches from famous tournaments such as the Six Nations League and the Rugby World Cup. For rugby league, we consider two different match categories: international matches from prominent tournaments such as the Super League and the Rugby League World Cup, and domestic matches played by British club teams.

Rugby league (union) is played between two teams of 13 (15) players. The rectangular playing field contains two try-lines across the width of the field, one on each side. These lines demarcate the in-goal areas. On the line, there are two posts with a crossbar. In rugby league (union), teams score four (five) points with a try, which happens when a team grounds the ball in the opposing team's in-goal area. Following a successful try, a team gets a conversion attempt, yielding two points if the team kicks the ball through the posts and over the crossbar from a chosen distance on the line perpendicular to the location where the try was scored. Teams score two (three) points if they kick a penalty between the posts, and one (three) by kicking the ball through the posts during game play. Matches consist of two 40-minute periods, separated by a 10-minute half-time break.

We obtained data for 2,475 rugby union, 2,306 international rugby league and 11,340 domestic rugby league matches that satisfy the data requirements outlined in Section 3.2.[8] The exclusion of matches with a zero half-time score difference reduces the samples to 2,338, 2,057 and 8,690 matches, respectively. Table 3.5 gives summary statistics. On average, the two teams together scored approximately 25 points in the first half, and 50 points in the whole match. At half-time, home teams on average led by 1 point (union) or 3 points (international league and domestic league). At full-time, the average score

[8]We scraped the data for rugby union matches from stats.espnscrum.com on 11 September 2018, for international rugby league matches from www.rugbyleagueproject.org on 7 November 2018, and for domestic rugby league matches from www.rugby-league.com on 5 October 2018. ESPN is primarily known as a sports TV channel, and their website offers extensive rugby union statistics. The Rugby League Project is a volunteer-run rugby statistics website, and rugby-league.com is the official website of the Rugby Football League.

Table 3.5: Summary statistics rugby

Panel A: Rugby union (1990-2018, N=2,338)							
	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	23.6	9.5	7	16	22	29	87
Total points at full-time	47.0	17.5	9	35	45	57	162
Score difference at half-time	0.5	12.3	−68	−7	1	8	81
Score difference at full-time	1.7	24.1	−152	−11	2	14	128
Home team wins match	0.55	0.50	0	0	1	1	1
Panel B: International Rugby League (1957-2017, N=2,057)							
	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	21.6	9.3	1	15	20	28	58
Total points at full-time	45.0	15.4	4	34	44	56	114
Score difference at half-time	3.1	12.9	−38	−6	4	12	52
Score difference at full-time	5.6	21.5	−74	−8	6	19	106
Home team wins match	0.61	0.49	0	0	1	1	1
Panel C: Domestic Rugby League (2006-2018, N=8,690)							
	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	24.8	9.7	2	18	24	30	84
Total points at full-time	51.5	16.1	6	40	50	62	144
Score difference at half-time	3.0	15.3	−68	−8	4	12	82
Score difference at full-time	5.9	27.0	−130	−12	6	22	144
Home team wins match	0.59	0.49	0	0	1	1	1

Notes: The table displays the summary statistics for rugby union, international rugby league and domestic rugby league matches where the half-time score difference was not zero. All definitions are as in Table 3.1.

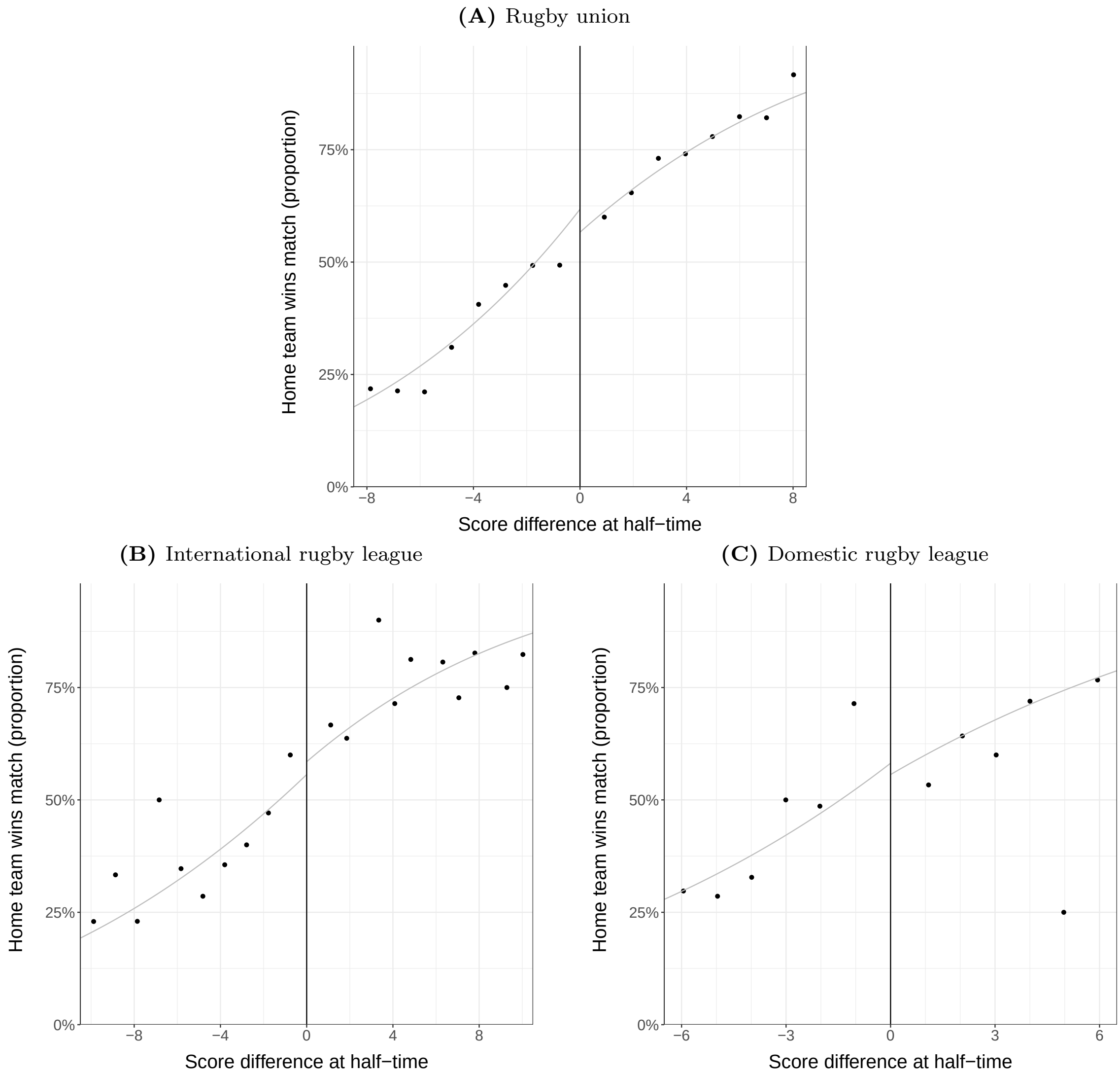
difference is 2 (union) or 6 (international league and domestic league). Home teams won approximately 60 percent of the matches.

3.5.2 Analysis and Results

Figure 3.3 shows that there is an approximately linear relation between the winning probability and the half-time score difference in rugby. In each sample, the winning probability increases at a rate of roughly four percentage points per point. All three samples exhibit some indication of a discontinuity in the winning probability at the half-time score difference of zero, but the sign of the visual discontinuity differs. For rugby union and domestic rugby league matches, the discontinuity suggests that trailing increases the chance of winning the match, whereas for international rugby league matches it suggests the opposite.[9]

[9] As a consequence of the scoring system, odd score differences are relatively rare in rugby league matches. This explains why some dots in Panel B and Panel C deviate sharply from the smooth curve. For example, the domestic rugby league sample includes only four matches where the half-time score difference between the home and away team was five, whereas there are 460 (609) matches where the difference was four (six).

Figure 3.3: Regression discontinuity plots for rugby



Notes: The figure shows the regression discontinuity plots for rugby union (Panel A), international rugby league (Panel B) and domestic rugby league (Panel C) matches. Definitions are as in Figure 3.1.

Table 3.6 shows the results for the main RDD. We find no convincing evidence that half-time trailing discontinuously affects the chance of ultimately winning the match. The estimated effect of trailing ranges from a 2.9 percentage point decrease (international rugby league) to a 6.5 percentage point increase (domestic rugby league). Notwithstanding these considerable effect sizes, all are statistically insignificant (all $p > 0.242$). Figure 3.8 in the Appendix shows the estimated treatment effects for a range of imposed alternative bandwidths. The rugby union and international rugby league results are not very sensitive. The estimated treatment effect for domestic rugby league matches increases considerably

Table 3.6: Results for rugby

	Rugby union	International rugby league	Domestic rugby league
Behind at half-time	0.033	−0.029	0.065
	(−0.112, 0.179)	(−0.194, 0.135)	(−0.044, 0.175)
Bandwidth	8.70	10.31	6.97
Total observations	2,338	2,057	8,690
Included observations	1,249	1,259	3,056

Notes: The table reports the estimated effect of being behind on the likelihood of winning for rugby union, international rugby league and domestic rugby league matches matches using a regression discontinuity design. Definitions are as in Table 3.2.

when we impose more restrictive bandwidths, but remains statistically insignificant at the five percent level. We cannot analyze the effect of being behind on a quarter-by-quarter basis, because rugby matches do not consist of quarters.

Last, we examine whether the skill difference between home and away teams is continuous through the cutoff. The results point out that there is no reason to doubt the validity of the RDDs: the discontinuity estimates are insignificant for all three samples (all $p > 0.157$).

In conclusion, and consistent with most of the previous analyses for Australian football and American football, rugby offers no compelling evidence that trailing at half-time improves the odds of winning.

3.6 Basketball

3.6.1 Description and Data

Thus far, we found little to no evidence that being behind improves the odds of winning in Australian football, American football and rugby. We now turn to basketball—the sport that is central in the study of BP—and consider four different samples. The first contains independently collected data for National Basketball Association (NBA) matches that took place in the same period as the NBA matches analyzed in BP. The second contains older and more recent NBA matches. The NBA is widely considered to be the premier basketball competition in the world, and pays the highest average salary of all the world's sports competitions.[10] Following BP, we also examine basketball matches of the National Collegiate Athletic Association (NCAA), the association that organizes the main intercollegiate competition in the US. Our fourth sample contains matches of the Women's National Basketball Association (WNBA), the women's counterpart to the NBA.

Basketball is played by two teams of five players each. The aim of the game is to score points by shooting a ball through the opposing team's hoop. Teams obtain two points by successfully throwing the ball through the hoop from the area inside the three-point arc, a semi-circle around the hoop, and three points by throwing the ball through the hoop from beyond the arc. After a foul, a team gets awarded one or more free throws, which are worth one point each. NBA and WNBA matches are played in four quarters of 12 and 10 minutes, separated by a 10-minute half-time break and two 2-minute breaks after the first and third quarter. NCAA matches are played in two 20-minute halves and have a 15-minute half-time break.

We obtained data for 35,921 NBA, 70,484 NCAA, and 4,666 WNBA matches that satisfy the criteria outlined in Section 3.2.[11] Approximately half of the NBA matches are from the period of 5 November 1993 to 1 March 2009 that was analyzed in BP. This subset, henceforth the "NBA BP" sample, contains 18,230 matches (BP's sample contains 18,060 matches). The sample of remaining NBA matches, henceforth the "NBA non-BP" sample, contains 17,691 matches played between 14 June 1987 and 20 June 1993 or between 2 March 2009 and 8 June 2018. The NCAA data cover the years 2006-2020. To have a clean out-of-sample test for the NCAA, we exclude matches that were played before 23 March 2009. This leaves 55,857 NCAA matches. The exclusion of matches with a zero half-time score difference reduces the four different samples to 17,535 (NBA BP), 17,001 (NBA non-BP), 53,770 (NCAA) and 4,499 (WNBA) matches.

Table 3.7 summarizes the data. On average, the two teams together scored around 200 (NBA) or 140 (NCAA and WNBA) points per match. At half-time, these averages

[10] Business Insider. 2015. The NBA is the highest-paying sports league in the world. Available from https://www.businessinsider.com/sports-leagues-top-salaries-2015-5 [Accessed: 7 July 2020].

[11] We scraped the NBA data from www.basketball-reference.com, a fan-edited basketball website, on 14 September 2018. We scraped the NCAA data from www.cbssports.com, the sports channel of the American TV network CBS, on 18 July 2020. We received the WNBA data from Michael Beuoy of www.inpredictable.com, a fan-edited prediction website, on 16 October 2018.

Table 3.7: Summary statistics basketball

Panel A: NBA BP (1993-2009, N=17,535)	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	97.1	12.4	55	89	97	105	152
Total points at full-time	193.0	20.0	121	179	193	206	286
Score difference at half-time	2.3	10.2	−39	−5	3	9	39
Score difference at full-time	3.6	13.3	−52	−6	5	12	65
Home team wins match	0.61	0.49	0	0	1	1	1
Panel B: NBA non-BP (1987-1993, 2009-2018, N=17,001)	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	103.1	12.8	58	95	103	111	174
Total points at full-time	205.0	20.8	133	191	204	219	320
Score difference at half-time	2.4	10.6	−41	−5	3	10	47
Score difference at full-time	3.9	13.7	−58	−6	5	13	68
Home team wins match	0.62	0.49	0	0	1	1	1
Panel C: NCAA (2009-2020, N=53,770)	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	65.5	11.5	22	57	65	73	146
Total points at full-time	139.3	19.4	65	126	139	152	241
Score difference at half-time	3.9	10.9	−40	−4	4	11	62
Score difference at full-time	6.9	15.7	−59	−4	7	16	104
Home team wins match	0.67	0.47	0	0	1	1	1
Panel D: WNBA (1997-2018, N=4,499)	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Total points at half-time	71.9	12.8	26	63	72	80	119
Total points at full-time	146.8	20.1	78	133	146	160	217
Score difference at half-time	1.9	9.8	−32	−5	2	9	45
Score difference at full-time	3.5	13.0	−45	−7	5	12	59
Home team wins match	0.61	0.49	0	0	1	1	1

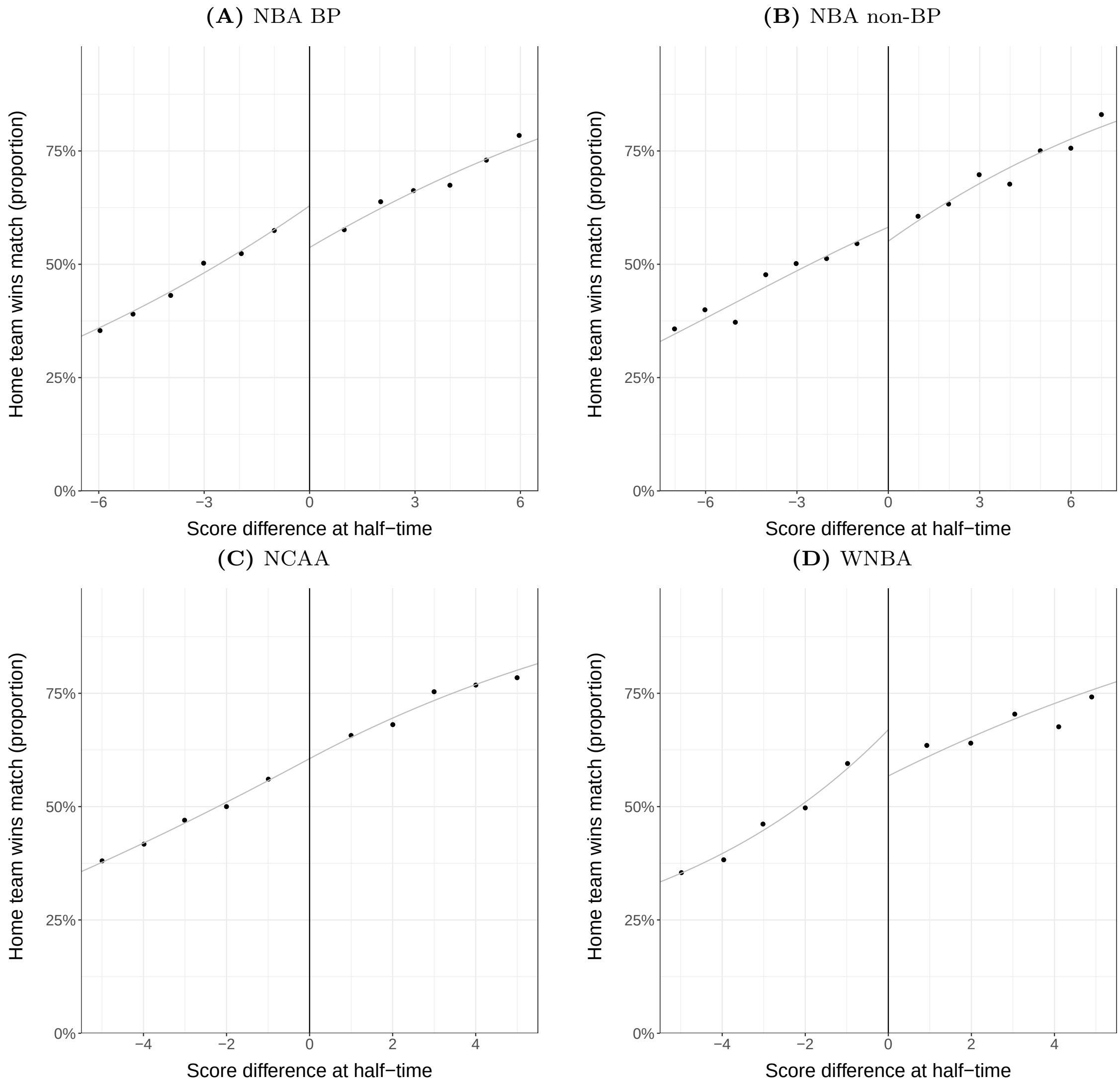
Notes: The table displays the summary statistics for NBA BP, NBA non-BP, NCAA and WNBA matches where the half-time score difference was not zero. Definitions are as in Table 3.1.

were approximately 100 and 70. The average score differences at half-time and full-time were around 2 and 4 (NBA and WNBA), or 4 and 7 (NCAA) points, respectively. Home teams won approximately 61 percent (NBA and WNBA) or 67 percent (NCAA) of the matches.

3.6.2 Analysis and Results

Figure 3.4 shows that the winning probability increases roughly linearly with the half-time score difference, at a rate of roughly four percentage points per point in all four samples. In line with the findings in BP we visually observe a negative discontinuity at a zero half-time score difference for the NBA BP sample, suggesting that marginally trailing at half-time increases the likelihood of winning the match. Visual discontinuities

Figure 3.4: Regression discontinuity plots for basketball



Notes: The figure shows the regression discontinuity plots for NBA BP (Panel A), NBA non-BP (Panel B), NCAA (Panel C) and WNBA (Panel D) matches. Definitions are as in Figure 3.1.

for the NBA non-BP sample and for the WNBA sample similarly suggest that there is a performance-enhancing effect of being behind. There is no indication of such an effect for NCAA matches.

Table 3.8, Panel A shows the results for the main RDD. For the NBA BP sample, we find that trailing improves the odds of winning by 8.3 percentage points ($p = 0.015$). For the same sample period, BP report an increase of 5.8 to 8.0 percentage points. Hence, our point estimate of the positive effect of trailing is even slightly higher.[12]

For the NBA non-BP, NCAA and WNBA samples, however, the effects of being behind are all statistically insignificant. The point estimates are 0.9 ($p = 0.788$), -0.1 ($p = 0.950$) and 6.0 percentage points ($p = 0.417$), respectively.[13,14] Figure 3.9 in the Appendix shows that these results are robust to imposing alternative bandwidths.

BP show that the effect of half-time trailing in NBA matches is stronger in the third quarter than in the fourth quarter. Because NCAA matches do not consist of quarters and because we do not have quarter-by-quarter scoring data for the WNBA, we can only conduct such an analysis for the two NBA samples. We exclude matches in which the competing teams scored the same number of points in the quarter of interest. Panels B and C show the results. Being behind at half-time increases the chance of winning the third quarter by 2.9 percentage points for NBA BP matches, but, in contrast to the results in BP, this effect is statistically insignificant ($p = 0.408$). In the NBA non-BP sample, the estimated treatment effect for the third quarter is negative and not significantly different from zero ($\tau = -0.027, p = 0.516$). The effect on winning the fourth quarter is statistically insignificant in both samples (NBA BP: $\tau = 0.018, p = 0.605$; NBA non-BP: $\tau = 0.027, p = 0.469$).

We also examine the effect of trailing after the third quarter on the probability of winning the match. This analysis includes matches in which the half-time score difference was zero, and excludes those in which the score difference after the third quarter was zero. Panel D shows that the treatment effect is insignificant in both the NBA BP ($\tau = 0.003, p = 0.935$) and the NBA non-BP sample ($\tau = -0.018, p = 0.574$), suggesting that trailing after the third quarter does not lead to better performance.

[12] The difference can be considered relatively small in the light of the somewhat different methodological approaches and the independently collected data. BP estimate the treatment effect with a standard logit model, for matches with a half-time score difference that falls within an ad hoc fixed bandwidth of ten points around the cutoff value of zero. If we conduct all analyses in the present chapter with their method, our conclusions remain unchanged.

[13] For the combination of the NBA BP and NBA non-BP data, the estimated effect is 5.0 percentage points ($p = 0.023$)

[14] For NCAA matches older than those in our sample (1999-2009), BP find that trailing at half-time increases the chance of winning, albeit by a smaller magnitude than for their NBA matches. If we do not exclude matches that were played before 23 March 2009—accepting that our sample period overlaps with that of BP—the estimated effect is 0.4 percentage points ($p = 0.821$).

Table 3.8: Results for basketball

	NBA BP	NBA non-BP	NCAA	WNBA
Panel A: Score difference at half-time, winning match				
Behind at half-time	0.083^{**}	0.009	−0.001	0.060
	(0.016, 0.150)	(−0.056, 0.074)	(−0.043, 0.040)	(−0.084, 0.204)
Bandwidth	6.06	7.32	5.25	5.67
Total observations	17,535	17,001	53,770	4,499
Included observations	7,938	8,513	19,183	1,792
Panel B: Score difference at half-time, winning third quarter				
Behind at half-time	0.029	−0.027		
	(−0.040, 0.098)	(−0.108, 0.054)		
Bandwidth	5.52	4.76		
Total observations	17,020	16,516		
Included observations	6,387	4,896		
Panel C: Score difference at half-time, winning fourth quarter				
Behind at half-time	0.018	0.027		
	(−0.050, 0.086)	(−0.046, 0.099)		
Bandwidth	5.67	5.76		
Total observations	17,032	16,499		
Included observations	6,418	6,014		
Panel D: Score difference third quarter, winning match				
Behind after third quarter	0.003	−0.018		
	(−0.063, 0.069)	(−0.081, 0.045)		
Bandwidth	5.66	6.20		
Total observations	17,274	16,780		
Included observations	8,615	9,543		

Notes: The table reports the estimated effect of being behind on the likelihood of winning for NBA BP, NBA non-BP, NCAA and WNBA matches using a regression discontinuity design. Definitions are as in Table 3.2.

The validity of our RDDs is not rejected by evidence against the continuity assumption: there is no significant discontinuity in the skill difference between home and away teams at the cutoff value of zero, neither at half-time nor after the third quarter (all $p > 0.225$).

In summary, we replicate the finding that trailing at half-time improves the odds of winning for NBA matches from the period analyzed in BP. There is, however, no evidence of such an effect for NBA matches outside of this sample period, and neither for NCAA matches nor for WNBA matches.

3.7 Meta-Analysis

Figure 3.5 summarizes the results of the previous sections for the main RDD. Trailing at half-time significantly improves the odds of winning in NBA matches from the period

Table 3.9: Power calculations

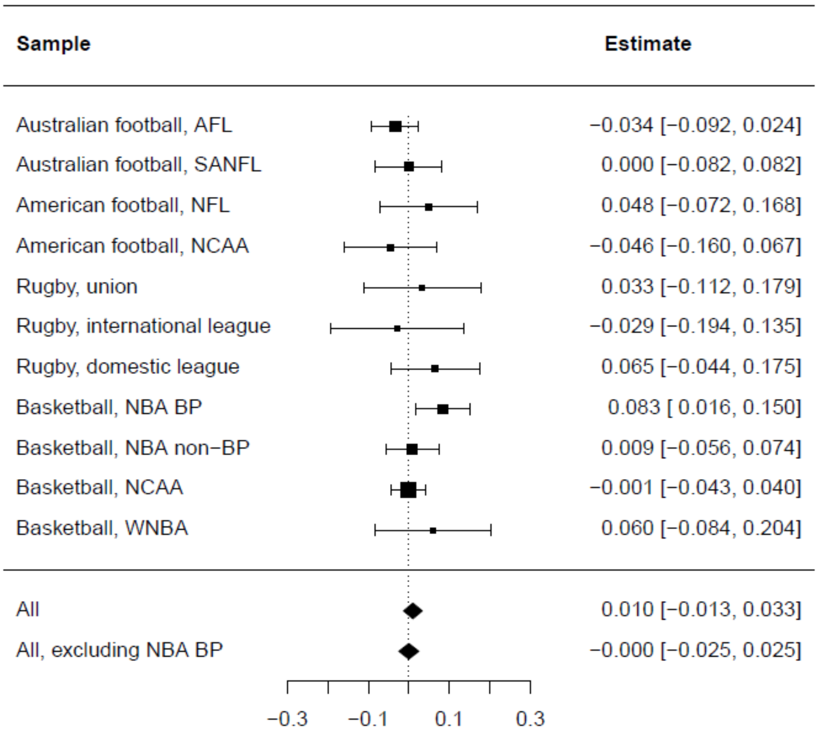
	Matches	Power
Australian football		
AFL	14,945	0.501
SANFL	6,622	0.284
American football		
NFL	10,590	0.158
NCAA	7,024	0.170
Rugby		
Rugby union	2,733	0.119
International rugby league	2,057	0.106
Domestic rugby league	8,690	0.180
Basketball		
NBA BP	17,535	0.400
NBA non-BP	17,001	0.420
NCAA	53,770	0.787
WNBA	4,499	0.124
Meta-analysis		
All	145,071	0.998
All, excluding NBA BP	127,536	0.996

Notes: This table presents the power calculations for both the main individual analyses and the meta-analyses. *Matches* is the number of matches in which the half-time score difference is not zero. *Power* is the probability of finding an estimate that is significant at the five percent level if the true effect is 0.058. The power of the individual samples is calculated with the approach proposed by Cattaneo et al. (2019), and the power of the meta-analysis is calculated with the analogical approach given in Jackson and Turner (2017).

analyzed in BP, but there is no evidence of such an effect in any of the other basketball samples or other sports that we have analyzed. To assess the informativeness of these null results, it is important to consider the statistical power of the underlying analyses. Statistical power refers to the likelihood of obtaining a significant estimate, under the assumption of a given true effect size. An analysis is generally considered to have sufficient power if it has an 80 percent probability of obtaining an estimate that is statistically significant at the five percent level.

To calculate the statistical power of each individual analysis, we use the approach proposed by Cattaneo et al. (2019), who developed a method to calculate the statistical power for the local-linear method proposed by Calonico et al. (2014). For the hypothetical true effect, we adopt the NBA estimates of BP, who find that half-time trailing improves the likelihood of winning by 5.8 to 8.0 percentage points. To be conservative, we assume that the true effect size is 0.058, the lower end of this range.

Figure 3.5: Meta-analysis for the effect of trailing at half-time on winning the match



Notes: The figure summarizes the main results for the individual samples, and shows the corresponding meta-analytic treatment effect estimates. The meta-analytic effects are estimated with the Paule-Mandel estimator (Paule and Mandel, 1989). The size of the squares represent the weights given to each sample. The lines (diamonds) represent the 95 percent confidence intervals for the individual analyses (meta-analyses). *Estimate* is the estimated effect of trailing at half-time on the chance of winning the match. Numbers in brackets represent 95 percent confidence intervals.

Table 3.9 shows the results of the power calculations. None of the individual analyses meets the 80 percent power benchmark. This is problematic if analyses are considered in isolation. Combined, however, these power statistics imply that the probability of finding insignificant estimates in all new samples—assuming a true effect of 0.058—is only 1.7 percent.

To synthesize the different results, we perform a meta-analysis. Because the true effect may differ across the different samples, we employ a random-effects meta-analytic model (Hedges and Vevea, 1998). The overall effect is the weighted average of the estimated treatment effects, where the weights are the inverse of the sum of the estimate's squared standard error and the estimated between-analysis variance. As recommended by Panityakul et al. (2013) and Veroniki et al. (2015), we estimate the between-analysis variance with the Paule-Mandel estimator (Paule and Mandel, 1989). The total number of matches underlying the meta-analysis is 145,071, or 59,788 if we only consider observations that are within the different bandwidths around the cutoff. Based on the analytical power calculation for meta-analyses as described in Jackson and Turner (2017), there is a 99.8 percent probability that our meta-analysis will detect a significant effect if the average true effect is 0.058. If we exclude the NBA BP sample, the power is 99.6 percent.

As shown in Figure 3.5, across all analyses the overall effect of being behind at half-time on the probability of winning the match is 1.0 percentage point, and statistically insignificant ($p = 0.398$).[15] If we leave out the analysis of the NBA BP sample and thus exclusively consider sports matches that have not been analyzed previously, the estimated overall effect size is economically and statistically indistinguishable from zero ($p = 0.996$). In both cases, the confidence intervals are relatively narrow.[16]

If performance improves only temporarily, an effect is more likely to emerge in the period directly following the half-time break. Figures 3.10 and 3.11 in the Appendix show the results of meta-analyses for the effect of half-time trailing on winning the third quarter and for the effect on winning the fourth quarter. The estimated overall effect of half-time trailing on winning the third quarter is 2.6 percentage points. This value is economically non-negligible, but statistically insignificant ($p = 0.208$). If we omit the corresponding analysis of the NBA BP sample, the coefficient is slightly higher, but it remains statistically insignificant ($\tau = 0.028, p = 0.283$). The meta-analytic estimates for the effect on winning the fourth quarter are negative and statistically insignificant: -0.5 percentage points ($p = 0.781$) for all analyses combined and -1.1 percentage points ($p = 0.557$) without the analysis of the NBA BP sample. In addition, Figure 3.12 shows that there is no meta-analytic evidence that trailing after the third quarter significantly affects the chance of winning the match, neither when the analysis of the NBA BP sample is included ($\tau = -0.007, p = 0.702$) nor when it is excluded ($\tau = -0.011, p = 0.615$).

[15] The estimated overall effect is 1.1 percentage points ($p = 0.321$) if we do not exclude NCAA basketball matches that were played before 23 March 2009.

[16] If we conduct meta-analyses for each sport separately, all four estimates are statistically insignificant (all $p > 0.199$).

In summary, our meta-analyses cannot reject the null hypothesis of no effect of marginally trailing on winning, and the confidence intervals suggest that the true effect, if existent at all, is likely relatively small.

3.8 Discussion and Conclusion

We extend Berger and Pope's (2011) analysis of whether marginally trailing improves the odds of winning in basketball to Australian football, American football and rugby. We find no clear evidence for these three sports: the estimated effects are sometimes positive and sometimes negative, and statistically almost always insignificant. We also revisit the phenomenon for basketball. For NBA matches from the period analyzed in BP we replicate the finding that half-time trailing improves the odds of winning, but for NBA matches from outside this period and for matches from the NCAA and WNBA we again obtain null results. Moreover, our high-powered meta-analyses across the different sports and competitions cannot reject the hypothesis of no effect of marginally trailing on winning, and the confidence intervals suggest that the true effect, if existent at all, is likely relatively small. This absence of supportive evidence is particularly informative in the light of BP's prior finding of a large positive effect and our sizable data sets (Abadie, 2020).

Australian football, American football and rugby are attractive sports for the analysis of interest, because of the large numbers of points that are generally scored. For reliably identifying a discontinuous effect of trailing on performance, it is important that the relation between the half-time score difference and the winning probability is approximately linear within a reasonable bandwidth around the cutoff value of a zero score difference. Australian football, American football and rugby satisfy this criterion, as demonstrated by the different regression discontinuity plots. Also, for the hypothesized psychological phenomenon to arise, it is important that the negative impact of trailing on the winning probability is limited, such that teams that are slightly behind still have a reasonable chance of winning. Otherwise, trailing by one or a few points likely discourages rather than motivates (Fershtman and Gneezy, 2011; Gill and Prowse, 2012). For Australian football, American football, and rugby, the relationship between the half-time score difference and the winning probability resembles the relationship for basketball, implying that the psychological effect of trailing should be similar across these four sports.

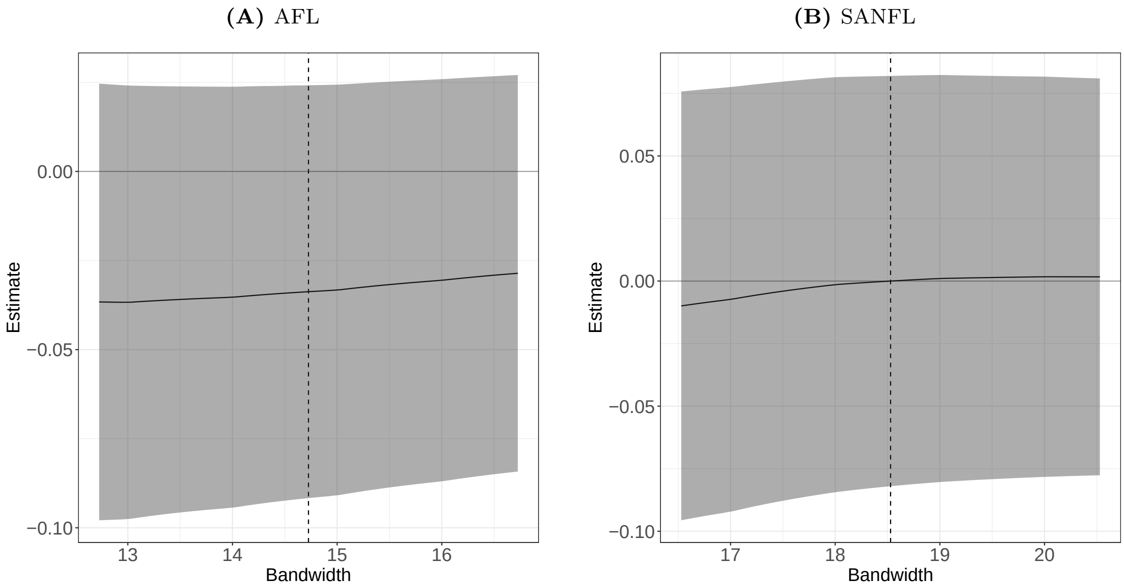
Our null results do not mean that being slightly behind in a competition does not or cannot have any systematic positive motivating effects. There is a robust literature that demonstrates that people who are slightly below their goal work harder than those who already reached it (see, for example, Heath et al., 1999; Pope and Simonsohn, 2011; Corgnet et al., 2015; Allen et al., 2016). In addition to the findings for basketball, BP present results from two laboratory experiments that show that this motivational effect also occurs during a competition. In a two-period button-pressing contest, subjects who were told after the first period that they are slightly behind worked harder in the second period than subjects who were told that they are far behind, tied, or slightly ahead. An important difference between sports matches and BP's laboratory task, however, is in the feedback that participants received. In the experiments, there was only one feedback moment, which precluded participants from responding to developments in the score difference after that. In sports matches, by contrast, players do get continuous feedback on the score difference. A disadvantage can turn into an advantage within mere seconds after a moment of reflection. Even if trailing is performance-enhancing and driving a turnaround in the short run, the effect may get lost in the chain of subsequent events, and the two teams' responses to these events.

Another potentially relevant difference is that professional athletes are highly experienced, whereas BP's laboratory subjects engaged in the button-pressing contest only once. If a leading team or subject realizes that their opponent will exert additional effort, they should anticipate this and adjust their own effort accordingly. Subjects in the laboratory may not realize that their trailing opponent will exert more effort, but can be expected to learn this if the game is repeated often enough. Therefore, the performance-enhancing effect of trailing may disappear with experience.

In the light of contest theory, our null results are not surprising. Contest theory considers situations in which agents have the opportunity to expend scarce resources to win prizes. A common prediction is that trailing by a considerable margin leads to further losing, because of the relatively weak incentive to exert effort (Harris and Vickers, 1987). Such a demotivating effect of trailing has been empirically confirmed in, for example, experiments (Dechenaux et al., 2015), tennis (Malueg and Yates, 2010; Page and Coates, 2017; Gauriot and Page, 2019) and political campaigns (Klumpp and Polborn, 2006). For infinitesimal score differences, however, contest theory predicts no material effect on effort and final outcomes.

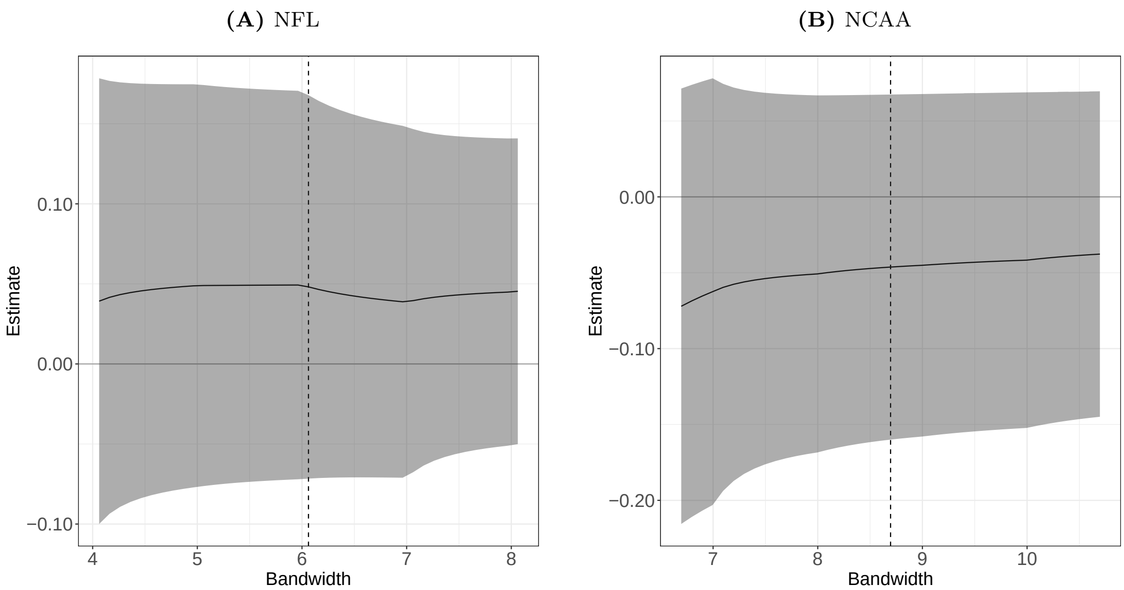
3.A Appendix

Figure 3.6: Bandwidth sensitivity Australian football



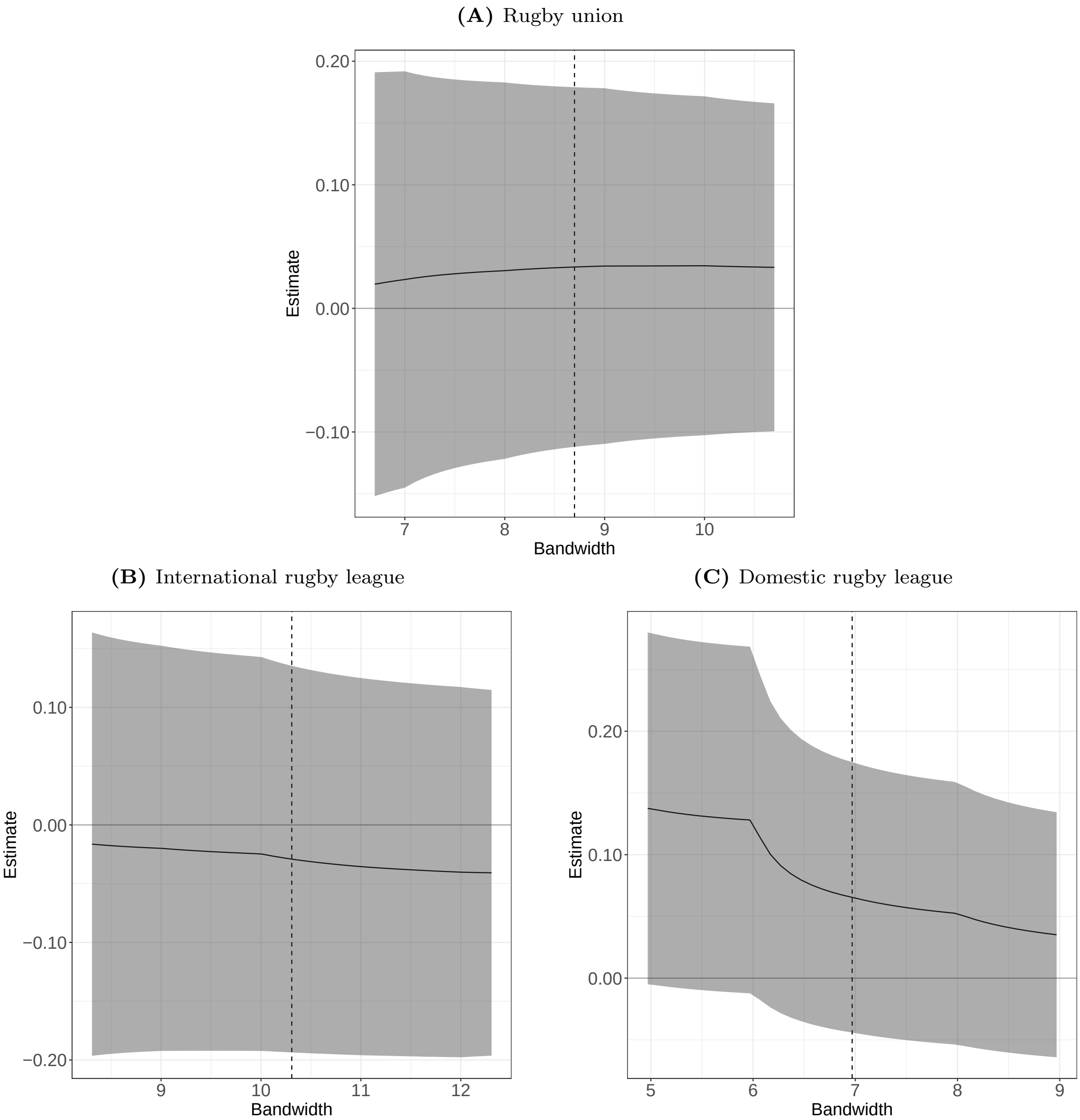
Notes: The figure shows the sensitivity of the main analyses for Australian football to a range of imposed alternative bandwidths. Compared to the bandwidth reported in Table 3.2, the smallest bandwidth is two points narrower and the largest is two points wider. The curves show the point estimates, the grey regions represent the 95 percent confidence intervals, and the dotted line shows the bandwidth of that minimizes the mean squared error.

Figure 3.7: Bandwidth sensitivity American football



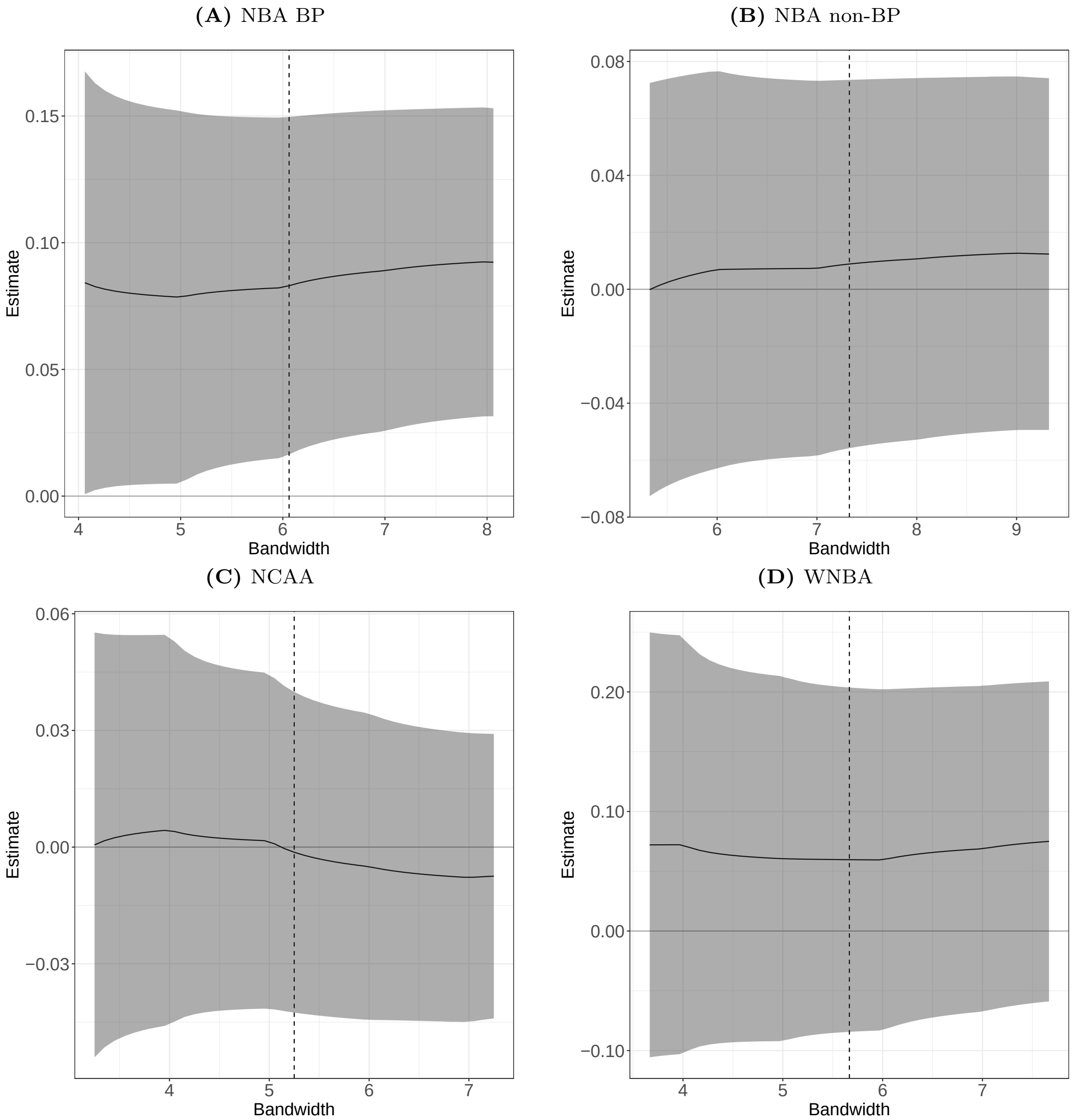
Notes: The figure shows the sensitivity of the main analyses for American football to a range of imposed alternative bandwidths. Definitions are as in Figure 3.6.

Figure 3.8: Bandwidth sensitivity rugby



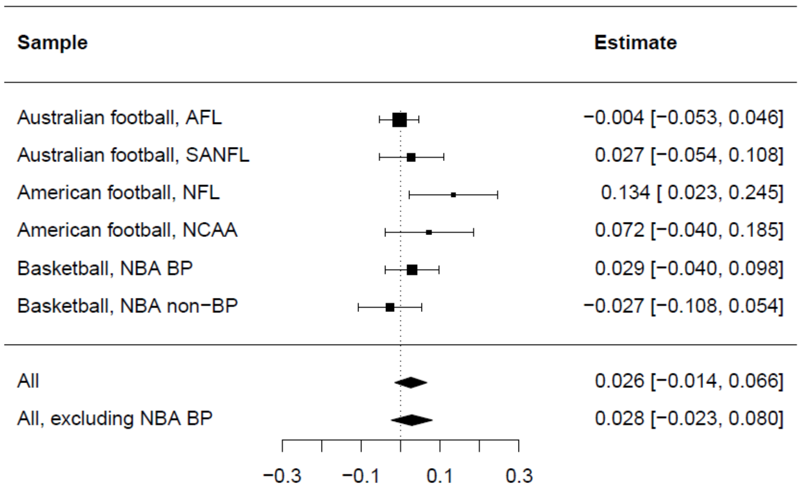
Notes: The figure shows the sensitivity of the analyses for rugby to a range of imposed alternative bandwidths. Definitions are as in Figure 3.6.

Figure 3.9: Bandwidth sensitivity basketball



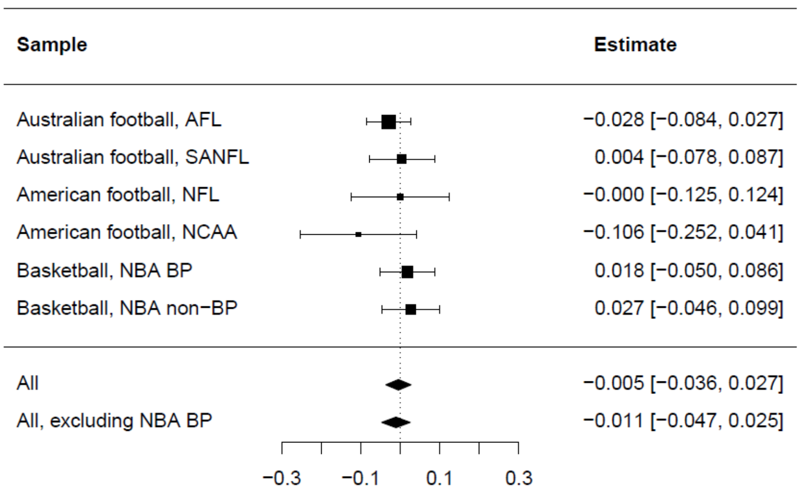
Notes: The figure shows the sensitivity of the main analyses for basketball to a range of imposed alternative bandwidths. Definitions are as in Figure 3.6.

Figure 3.10: Meta-analysis for the effect of trailing at half-time on winning the third quarter



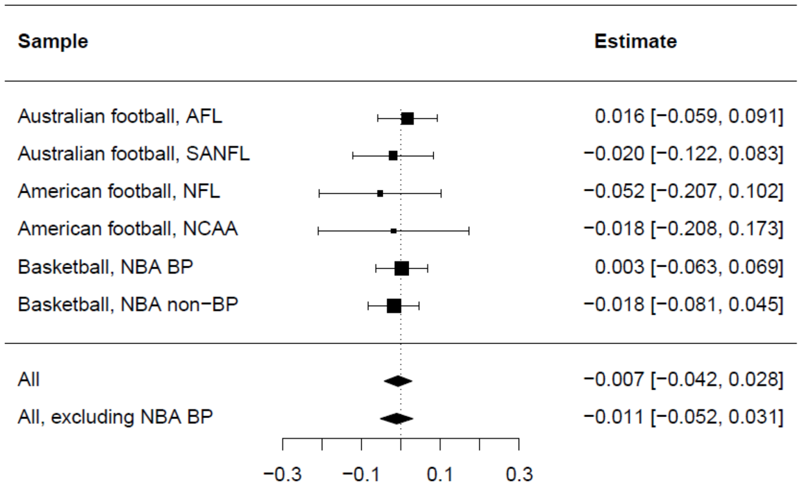
Notes: The figure summarizes the results of the analyses of the effect of trailing at half-time on winning the third quarter, both for the individual samples and for the meta-analyses. *Estimate* is the estimated effect of trailing at half-time on the chance of winning the third quarter. All definitions are as in Figure 3.5.

Figure 3.11: Meta-analysis for the effect of trailing at half-time on winning the fourth quarter



Notes: The figure summarizes the results of the analyses of the effect of trailing at half-time on winning the fourth quarter, both for the individual samples and for the meta-analyses. *Estimate* is the estimated effect of trailing at half-time on the chance of winning the fourth quarter. All definitions are as in Figure 3.5.

Figure 3.12: Meta-analysis for the effect of trailing after the third quarter on winning the match



Notes: The figure summarizes the results of the analyses of the effect of trailing after the third quarter on winning the match, both for the individual samples and for the meta-analyses. *Estimate* is the estimated effect of trailing after the third quarter on the chance of winning the match. All definitions are as in Figure 3.5.

Chapter 4

Discounts Shift the Demand Curve for Life-Saving Medications[1]

4.1 Introduction

Pharmaceutical pricing in the US is highly opaque, and many patients who arrive at a pharmacy to pick up medications prescribed by their doctor abandon them when they learn the cost. To increase the likelihood that patients purchase their medicine, pharmaceutical manufacturers have introduced copay assistance: discounts that lower patients' out-of-pockets costs. Apart from lowering the out-of-pocket costs, discounts may also instill on patients the impression they are getting a good deal, thus generating transaction utility—pleasure associated with the quality of a deal (Thaler, 1985). As a consequence, discounts may not only cause a movement along, but also a shift of, the demand curve.

We exploit a unique data set containing transaction data from approximately 85 percent of all US pharmacies to estimate the causal effect of discounts on patients' propensity to purchase their medicine. Consistent with transaction utility, we find that discounts shift the demand curve: patients who receive copay assistance are more likely to pick-up their medicine *at a given final price* than otherwise identical patients who do not.

This chapter is structured as follows. Section 4.2 discusses the pricing of medications and shows how prices are jointly borne by manufacturers, insurers, and patients.

[1]This chapter is based on joint work with Matthew Jordan, Nicholas Adolph and Shane Frederick.

Section 4.3 discusses the transaction utility introduced by copay assistance. Section 4.4 describe our data source and presents a nearest-neighbor matching analysis of first-time purchases on discounts. Section 4.5 presents a within-person analysis of the effect of discounts on repeated purchases. Finally, Section 4.6 provides a general discussion of our findings and their relevance to consumer decision making and policy designers.

4.2 Pricing Medications

In the United States, most citizens who have health insurance receive it through their employer or through government programs like Medicare and Medicaid (Berchick et al., 2018). Health insurers and pharmacy benefit managers (or PBMs) dictate patients' out-of-pocket costs for medication through copays (a flat dollar amount paid per drug), coinsurance (a percentage of the drug's cost paid by the patient), and/or deductibles (a set amount of out-of-pocket expenditure before an insurer will contribute to further expenses) (Daniel and Bornstein, 2019). Patients attempting to fill a prescription typically do not know their out-of-pocket costs until the pharmacist verifies their insurance benefit, and it is only then that patients learn what it will cost them to acquire their medication. Upon receiving the news, many patients forego the medication they were prescribed.

Drug manufacturers are typically aware of patients' benefit design landscapes and set drug prices accordingly. In addition to patients' price elasticity, the list price set by a manufacturer is informed by features like the product's therapeutic area, novelty, and clinical differentiation, expected payer coverage, and other market and benefits considerations. These processes mean that patients may not always be aware of the true list price of a drug.

Even when patients can know the list price and patient cost-burden is tied directly to that list price, manufacturers may be incentivized to set high prices. One such incentive comes from coinsurance benefit structures, where the patient pays a percentage of the drug's list price, typically 20 to 40 percent of the price set by the manufacturer. Coinsurance benefits can induce manufacturers to reduce the list price of the drug to make it affordable for patients. However, issuing post-coinsurance discounts (i.e., copay assistance) directly to patients can make it economical for manufacturers to increase the list price of a drug when the increase in coinsurance reimbursement is greater than the direct-to-patient price assistance.

Furthermore, the relationship between manufacturers, PBMs, and health plans complicates the drug pricing equation, creating a number of conflicts of interest (Dafny et al., 2017; King et al., 2019; Royce et al., 2019; Lee, 2020). Manufacturers and PBMs negotiate for rebates (paid by the manufacturer to the PBM) in exchange for preferential formulary placement (resulting in lower cost-sharing for patients). PBMs—who negotiate for rebates on behalf of health plans—have a financial incentive to prefer high list prices with high rebates paid to the PBM and plan, as they often receive a percentage of the rebate. For example, if Drug A is priced at $500 with a 25 percent rebate and Drug B is priced at $750 with a 50 percent rebate, the cost to consumers will be the same ($375). However, if the PBM receives a fraction of the rebate, they would prefer the drug with the higher list price (and, correspondingly higher rebate), in this case Drug B. This dynamic plays a role in price setting, and further incentivizes manufacturers to set a higher price and provide copay assistance, via mechanisms like copay assistance, to those patients exposed to the drug's list price.

Although a clear picture of the alignment of incentives across parties like insurers, PBMs, and patients is fundamental for policy-makers, our analysis will only deal with patient behaviors, namely drug abandonment and adherence. While important, questions around various pricing incentives, conflicts of interest, and efficiency are beyond the scope of this work, which is exclusively aimed at measuring the effects of discounts on patients' adherence to prescribed drug regimens.

Many drug manufacturers offer copay assistance programs to help offset the out-of-pocket cost of prescription medications (Yezefski et al., 2018). We focus here on copay assistance: manufacturer-issued coupons that reduce patient out-of-pocket expense. Most of these programs include a target copayment (the minimum amount patients pay) and a maximum benefit (a cap on how much the manufacturer can contribute). Only commercially-insured patients may legally use copay assistance programs, although there are mechanisms through which manufacturers can offer programs for government-insured patients to access drugs for free. We exclude all government-insured patients from our analysis.

4.3 Transaction Utility

The perceived quality of a deal can affect demand for a good beyond the direct effect of the price (Thaler, 1985, 1999). For instance, if contemplating the purchase of a product priced at $119, you might be less inclined to do so if you noticed the sign "Was $82." Although historical price might rationally affect quality inferences (Rao and Monroe, 1989) or beliefs about what that good would cost at some other place or some other time, contemplation that someone else once paid less may dissuade purchase via "bad deal aversion" (Isoni, 2011; Weaver and Frederick, 2012), which draws on Thaler's (1985; 1999) work on mental accounting.

Thaler's framework conceptualizes purchasing decisions as being jointly determined by acquisition utility $v(a)$—the utility of the good $v(p)$ minus the disutility of the cost $v(c)$—and transaction utility d—the perceived quality of the deal. Classical consumer theory postulates that consumption decisions solely depend on acquisition utility, and that consumers will only purchase if $v(p) > v(c)$. The transaction utility model, by contrast, suggest that consumers additionally receive utility from the quality of the deal d they are getting, and thus suggests that consumers evaluate a slightly modified inequality, $v(p) + d > v(c)$, to decide whether or not to purchase a good. Transaction utility predicts that a given price will generate less demand when construed as a surcharge (as in the illustrative, though rare example above) and more demand in the much more common cases where a price is perceived as being discounted.

Although these predictions draw support from hypothetical choices in many lab studies (e.g. Berkowitz and Walton, 1980; Moore and Olshavsky, 1989; Grewal et al., 1998; Campbell, 1999; Sinha and Smith, 2000; Muehlbacher et al., 2015; Huang, 2018), evidence from real purchase decisions is scarce, in part because bigger discounts generally entail lower prices as well as better deals. Distinguishing movements along the demand curve (the effect of the current price) from shifts of the demand curve itself (the effect of transaction utility) requires two features that are typically unobserved in non-experimental data. First, to bypass estimating full demand curves, a given product must be sold for the same final price with and without discount. Second, in addition to observing aggregate demand for a product at a given price, one must know whether an individual decides to not purchase a product, which is typically unobserved in sales data.

Our pharmaceutical transaction data set, which we describe in detail below, allows us to overcome both of these issues because we observe situations in which final prices are held constant while the presence of a discount varies and, notably, we know when a purchase is intended (indeed, prescribed) but not made. The introduction of copay assistance moves d from zero to a non-zero value. When d is positive, patients will be more likely to pick up their drugs, even if $v(p)$ and $v(c)$ remain constant. If d is positive, patients experience transaction utility, and the demand curve shifts outwards.

4.4 First-Time Purchase Decisions

Patients first enter our data set when they are initially prescribed a new medication. This new prescription is sent to the patient's pharmacy, filled, and either paid for or abandoned by the patient. If a patient abandons their prescription, they leave it unclaimed at the pharmacy. In the current section, we estimate whether the demand curve for first-time prescriptions shifts outwards as a result of transaction utility introduced by copay assistance. In Section 4.5, we examine repeated purchase decisions by exploiting within-patient variation in copay assistance availability.

4.4.1 Data Source

Throughout our analyses, we exploit a proprietary pharmaceutical transaction data set; an anonymized patient sample data set that captures longitudinal pharmacy (both retail and specialty) claims. This database captures prescription transactions at approximately 85 percent of all US retail pharmacies. It is made up of multiple sources, including national and regional chains, independent pharmacies, and a switch house for a comprehensive view into all types of retailers across all geographies. The complete data, from which our data is only a subset, is aggregated and maintained for many reasons, including to track patient behavior, copay assistance, and patient copay (including primary and final out-of-pocket cost) over time. In addition to capturing adjudicated (paid) prescriptions, this data source captures claims that are rejected by health insurers ("rejections") as well as claims that are approved by health insurers but not paid for by patients ("abandoned" claims). Thus, this data source captures all of the cost burdens and how they are divided across patients, insurers, and manufactures for each transaction. The longitudinal nature of the data set allows us to follow patients from the time they are first prescribed a new medication through each subsequent transaction they have at the pharmacy when their prescriptions are refilled.

We focus our analyses on transactions that took place between 2016 and 2019, involving eight drugs across three therapeutic areas: (1) inhalers for patients with asthma/COPD, (2) blood thinners for patients at risk of heart attack, stroke, pulmonary embolism, etc., and (3) non-insulin treatments for those with type 2 diabetes. To ensure the quality of our data, we exclude all patients who do not receive copay assistance but whose initial and final costs do not match, and who therefore must have access to other means of assistance. We furthermore exclude transactions for which the final costs are either equal to zero or below the target copayment (for reasons we detail later).[2] We furthermore exclude patients older than 65 because they receive government insurance that prohibits commercial copay assistance. The final data set for these eight drugs contains transactions from 1,616,016 unique patients. See Table 4.1 for descriptive statistics.

4.4.2 Identification Strategy

Our goal is to estimate the causal effect of a binary treatment—receipt of copay assistance—on the probability that a patient abandons their prescription while holding the final price constant. Because both the price of the drug and personal characteristics differ between patients who do and do not receive copay assistance, we use a nearest-neighbor matching estimator to obtain the effect (Abadie and Imbens, 2011). Such matching estimators aim to find, for each recipient of copay assistance, a non-recipient with the exact same prescription drug, zip code of prescribing physician, insurance coverage and specification, and the closest possible age, week of pick-up, and (post-discount) drug price.

[2]Our conclusions do not materially change if we do not exclude transactions with a price either equal to zero or below the target copay.

Table 4.1: Descriptive Statistics

Years	2016 - 2019
Observations	6,999,641
First encounters	1,736,070
Patients	1,616,016
Fraction female	57.5%
Age (mean)	48.5
Pick-up rate	89.0%
Copay rate	9.9%
Out-of-pocket price (median)	$30
Discount (median)	$50

Notes: The table displays the descriptive statistics. *Years* is the time period covered by the sample. *Observations* is the sum of the total number of purchases and the total number of abandonments. *First encounters* is the number of first-time purchases. *Patients* is the total number of patients. One patient can have multiple first encounters if she is prescribed more than one drug. *Fraction female* is the fraction of patients that is female. *Age (mean)* is the average patient's age. *Pick-up rate* is the fraction of prescriptions that is actually purchased. *Copay rate* is the fraction of customers that have copay available to them. *Out-of-pocket price (median)* is the median price that patients actually pay. *Discount (median)* is the median discount copay users receive.

Compared to standard OLS regressions, this procedure has the advantage of not assuming a functional form for the relation between the outcome variable and the covariates, nor does it extrapolate over areas of uncommon support in the observable characteristics.

More formally, let T_i denote the treatment variable that takes the value of 1 if patient i received copay assistance, and 0 otherwise. Let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes for patient i in treatment (patient received copay assistance) and control (patient did not receive copay assistance). The potential outcome takes a value of 1 in case of a purchase, and a value of 0 in case of abandonment. We are interested in the causal effect $Y_i(1)$ - $Y_i(0)$. The fundamental problem of causal inference, however, is that we only observe one value for each patient:

$$Y_i = T_i \times Y_i(1) + (1 - T_i) \times Y_i(0) \tag{4.1}$$

For patient i who receives copay assistance, the potential outcome $Y_i(0)$—the purchase decision without copay assistance—is unobserved. The nearest-neighbor matching procedure imputes this missing outcome by the decision of a patient who does not receive copay assistance but is otherwise observably identical. Let X_i denote the vector of observable covariates including drug, age, zip code of prescribing physician, week of pick-up (to account for annual benefits seasonality), drug, and (post-discount) price. A suitable match for patient i will be a non-recipient whose observable covariates are maximally similar to patient i's, such that the only observable difference is the copay assistance status. The prices of the assistance-recipients and matched non-recipients must lay within 0.1 standard deviation in order to be included as a match. We apply Abadie-Imbens bias correction to address remaining post-matching heterogeneity.[3] We use matching with replacement, allowing each untreated unit to be used as a match more than once. The missing potential outcomes for treated patients are thus given by

$$\hat{Y}_i(0) = \sum_{j=1}^{N_0} (1 - T_j) W(i,j) Y_j(0) \tag{4.2}$$

$W(i,j)$ is the weight given to observation j for computing patient i's purchase probability in the absence of copay assistance, and N_1 and N_0 denote the number of treated and untreated observations. Nearest-neighbor matching implies that we match treated individual i to the closest untreated individual:

$$W(i,j) = 1 \quad \text{if} \quad j = \arg\min_j \|X_i - X_j\| \tag{4.3}$$

In the case of ties, both observations receive a weight of 0.5. We estimate the average treatment effect as:

$$\tau = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_i - \sum_{j=1}^{N_0} W(i,j) Y_j \right) \tag{4.4}$$

[3]We also fit a number of models to check the robustness of our treatment effect estimate, by, for example, varying how we define closeness of time (same calendar week vs nearby day of the year) or price. None of these choices, either loosening or tightening matching restrictions, qualitatively change our results.

Table 4.2: Effect of copay assistance on pick-up rates matching analysis

Copay	0.047***
	(0.001)
Total observations	1,736,070
Effective observations	32,466

Notes: The table reports the estimated effect of the likelihood that a patient purchases their newly prescribed medicine on the patient's copay assistance status. Treatment effects are estimated using a nearest-neighbor matching method with bias correction (Abadie and Imbens, 2011). *Total observations* is the total number of first-time purchases. *Effective observations* is the number of matched observations. Numbers in parentheses represent the standard errors. Asterisks denote significance at the 0.01 (***), 0.05 (**) and 0.1 (* level.

Inferring that the effects estimated using this matching procedure are, in fact, causal effects, requires that copay assistance is independent of $Y_i(0)$ and $Y_i(1)$ conditional on $X_i = x$.

4.4.3 Results

As Table 4.2 shows, when we match patients on age, zip code of prescribing physician, week of pick-up, drug, insurance plan, and post-discount price, we find that patients who had access to copay assistance were 4.7 percentage points more likely to pick up their medication than those who did not. To place this 4.7 percentage point decrease in abandonment in context, in our total sample, 11 percent of people abandon their medications. See Table 4.8 in the Appendix for additional matching specifications that show the robustness of this effect across a variety of matching criteria.

4.4.4 Discussion

Recall that the classical view of prescription pick-up rates says that discounts matter only insofar as they reduce price; holding final out-of-pocket cost constant, discounts should leave demand unaltered. This contrasts with the transaction utility view, in which discounts should increase demand even holding final-out-of-pocket cost equal. The pattern of data we observe in our analysis of abandonment is suggestive of a transaction utility effect. When we match individuals with the same insurance, who are purchasing the same drug, around the same time, after receiving a prescription from a physician in the same zip code, and who pay a similar out-of-pocket price, we find patients with a discount are less likely to abandon their drug at the pharmacy. In other words, on the assumption that the value of the drug $v(p)$ and utility cost $v(c)$ are equal for patients with and without copay assistance, the transaction utility d must be positive in order to account for the increased pick-up rates among those who receive copay assistance. The key question then becomes: what reason do we have to believe that $v(p)$ and $v(c)$ are equal for patients with and without copay assistance? Might it be the case, for example, that wealthier or healthier patients are less likely to receive copay assistance or simply have a lower disutility of $v(c)$?

Part of the answer to these questions comes from our matching process, which is intended to create a *prima facie* plausible counterfactual on the basis of observables. Individuals who have the same insurance and who have been prescribed the same drug, around the same time, by a physician in the same zip code, at a similar out-of-pocket price are similar in many relevant ways. The confounding would have to come about through a patient-level association between receiving copay assistance and either $v(p)$, $v(c)$, or both. This could come about if drug manufacturers know something about patients that is unobserved in our data and target a subset of patients with copay assistance on the basis of those unobservables; this is possible, but unlikely given that our data source usually is the drug manufacturer's data source. In other words, the space of possible confounders is far more constrained in our setting than is typical of causal inference problems, because would-be unobserved confounding in treatment would have to originate from covariates we can control for. Nevertheless, one might be tempted to argue it is possible that copay assistance varies non-randomly across patient values of $v(p)$, $v(c)$, or both.

To address concerns that our matching procedure does not fully rule out heterogeneity in patient-level values of $v(p)$ and $v(c)$, we take advantage of the longitudinal nature of our data set and fit patient-price fixed effects models, functionally creating a within-subjects experiment where copay assistance, but not out-of-pocket cost, varies over time within an individual.

4.5 Repeated Purchase Decisions

The previous section dealt with whether or not a patient is more likely to start their prescription regimen when they receive a discount on their medication compared to otherwise identical patients who pay the same final price but receive no discount. In this section we address an arguably more difficult patient behavior problem: treatment adherence, a patient's ability to continue their treatment regimen for as long as they are being prescribed their medication. We exploit within-patient variation in the availability of copay assistance to estimate the effect of discounts on the likelihood that a given person purchases a given drug at a given price.

4.5.1 Data and Identification Strategy

We use the same data source for our adherence analysis as we did for our abandonment analysis. However, here we look at all pharmacy visits (in the year following first approval), rather than only examining first-time pick-ups for a given patient-drug combination. To maximally control for unobserved heterogeneity, we restrict our analysis to patient-price combinations, meaning that we only examine situations in which a given patient has at least one transaction with and without copay assistance at a given price. This restriction leaves us with 198,358 drug transactions by 29,049 patients. Six of our eight drugs are represented in this restricted data set.

We estimate a patient-price fixed effects model to investigate whether discounts raise demand at a given price. Furthermore, we examine whether the presence of a discount interacts with the size of the discount, the final price, or the number of prior transactions/encounters a given patient to predict pick-up rates.

4.5.2 Results

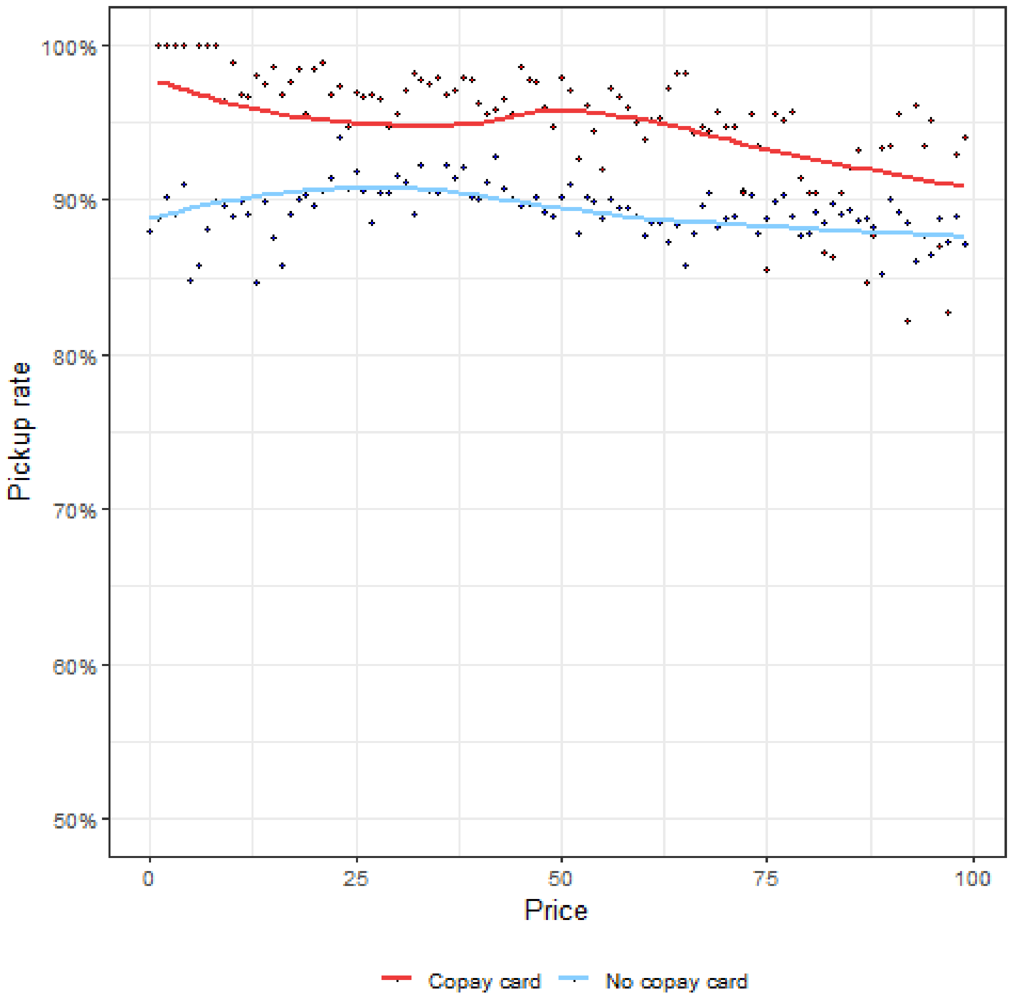
Recall that we are considering two theories of consumer choice—classical consumer theory and transaction utility theory—and these theories make diverging predictions about whether the presence of a discount should increase patients' propensity to purchase their medicine over and above how much it moves patients along the demand curve. We summarize the effect of receiving copay assistance (along with size of discount, final price, and number of encounters), in Figure 4.1 and Table 4.3 below. Figure 4.1 plots the raw pick-up rates for patients with and without access to copay assistance as a function of final out-of-pocket price (see Figure 4.2 for a visualization of estimated effects). Model 1 in Table 4.3 shows that a given patient is 1.6 percentage points ($p < 0.001$) more likely to pick up their medication when they have copay assistance compared to when that same patient paid the same out-of-pocket price for the same drug without copay assistance. Specifications 2-4 show weak or non-existent evidence of interactions between copay assistance and the size of the discount, final price, or encounter number.

Model 2 in Table 4.3 shows that we find no evidence of a dose dependent effect of discount size on pick-up rates; this is, perhaps, surprising as it would suggest that the presence of a discount outweighs the size of the discount in their effects on patients' medication demand. In the Appendix, we present a patient-price fixed effects analysis among patients receiving copay assistance to estimate the effect of discount size on pick-up rates, holding copay assistance status constant. That analysis finds that for every \$100 increase in discount, patients are 0.2 percentage points more likely to pick-up their medication, a statistically significant but small effect of discount size.

4.5.3 Robustness Checks

The basic result presented above includes all patients in our data set that experienced the target within-subjects treatment we sought. However, there are additional restrictions we can place on our inclusion criteria that make it possible to explore how a number

Figure 4.1: Demand curves for copay users and non-copay users



Notes: The figure shows the demand curves for copay users (red) and non-copay users (blue). It reports the average medicine purchase rates for all one-dollar bins of final out-of-pocket prices between $0 and $100. The smoothed curves represent loess regressions.

of potential confounders and psychologically relevant factors alter the effect of copay assistance.

Price paid is above target price.—Many pharmaceutical manufacturers explicitly market a desired patient out-of-pocket cost for their medications on their marketing website or in advertisements, thus providing (some) patients with *ex ante* knowledge of drug prices. In order to remove the possibility that *ex ante* price knowledge drives the effect of copay discounts on demand, we examine only those patients who paid an out-of-pocket price above the advertised target price. As the results of Table 4.4 show, restricting our analysis to patients who paid more than the target price only increases our estimate of the effect

Table 4.3: Effect of copay assistance on pick-up rates per patient-price combination

	Model 1	Model 2	Model 3	Model 4
Copay	0.016***	0.015***	0.016***	0.013***
	(0.001)	(0.001)	(0.001)	(0.002)
Copay x Discount		0.00001		
		(0.00001)		
Copay x Final price			−0.00000	
			(0.00002)	
Encounter number				0.0002
				(0.0002)
Copay x Encounter number				0.0005
				(0.0003)
Patient-price fixed effects	Yes	Yes	Yes	Yes
Observations	198,358	198,358	198,358	198,358
Adjusted R^2	0.140	0.140	0.140	0.140

Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance status. *Copay* is a dummy variable that takes the value of 1 if the patient has copay assistance available *Discount* is the copay discount, *Final price* is the out-of-pocket costs, and *Encounter number* is the number of prior transactions the patient has had. The regression specifications include patient-price fixed effects. Definitions are as in Table 4.2.

of copay assistance on demand. As Model 1 in Table 4.4 shows, restricting to instances in which patients pay more than the target price reveals a copay assistance effect of 2.2 percentage points ($p < 0.001$).

Price paid is equal to target price.—One point we have repeated throughout is that patients are unlikely to know their out-of-pocket cost until they get to the pharmacy counter. While this is typically true, the only cases in which it is possible for patients not to be surprised are when the out-of-pocket price is equal to the manufacturer target price. In other words, while many patients will not know the target price, some may know it, which could reduce the effect of copay assistance on patients' demand if discounts produce more transaction utility when they are a surprise rather than priced in from before the purchase decision.

As Table 4.5 shows, restricting our analysis to only those patients who paid an out-of-pocket price equal to the manufacturer target price does not qualitatively alter the effect

Table 4.4: Effect of copay assistance status on pick-up rates per patient-price combination, price above target

	Model 1	Model 2	Model 3	Model 4
Copay	0.022***	0.023***	0.028***	0.002
	(0.004)	(0.005)	(0.005)	(0.008)
Copay x Discount		−0.00000		
		(0.00003)		
Copay x Final price			−0.0001*	
			(0.00003)	
Encounter number				0.002
				(0.001)
Copay x Encounter number				0.004***
				(0.001)
Patient fixed effects	Yes	Yes	Yes	Yes
Observations	21,687	21,687	21,687	21,687
Adjusted R^2	0.201	0.201	0.201	0.202

Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance status for all observations for which the out-of-pocket costs are above the target price. Definitions are as in Table 4.3.

of copay assistance on patient's pick-up likelihood. In this restricted data set, copay assistance induces a 1.5 percentage point ($p < 0.001$) increase in the likelihood patients purchase their medicine, compared to the 1.6 percentage point ($p < 0.001$) bump we observed across all out-of-pocket/target price combinations presented in Table 4.3 above.

Patient-price-year fixed effects.—Thus far, we have put to one side the role of time on health and health expenses. In particular, a patient's health condition may change over time, which could affect both their choice of health plan (and hence their deductible and copay schedule) and their propensity to purchase their medicine. Or put more generally, there may be time-varying unobservables that either 1) create a spurious correlation between receipt of copay assistance and prescription pick-up rates, or 2) act as a common cause of an increase in how likely patients are to receive copay assistance and to adhere to prescription regimens. To eliminate the possibility that health changes over time may covary with health plan choices to create time-varying confounds, we fit a model with patient-price combinations that occur in a given year. Because patients cannot

Table 4.5: Effect of copay assistance status on pick-up rates per patient-price combination, price equal to target

	Model 1	Model 2	Model 3	Model 4
Copay	0.015***	0.014***	−0.004	0.013***
	(0.001)	(0.002)	(0.003)	(0.002)
Copay x Discount		0.00001		
		(0.00001)		
Copay x Final price			0.001***	
			(0.0002)	
Encounter number				0.0002
				(0.0002)
Copay x Encounter number				0.0002
				(0.0003)
Patient-price fixed effects	Yes	Yes	Yes	Yes
Observations	176,671	176,671	176,671	176,671
Adjusted R^2	0.114	0.114	0.114	0.114

Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance status for all observations for which the out-of-pocket costs are equal to the target price. Definitions are as in Table 4.3.

change health plans mid-year (except in rare circumstances, mostly unrelated to health like a change in marital status), this patient-price-year fixed effects model removes the temporal components that could create covariation between health and health expenses. As Table 4.6 shows, our patient-price-year fixed effects model estimates a very similar increase in demand, namely 1.4 percentage points ($p < 0.001$), compared to the 1.6 percentage points estimated by models that do not include year fixed effects.

4.5.4 Gaining and Losing Discounts

Our patient-price fixed effects models analyze pick-up rates for a given patient paying a given price for a drug, with and without a discount. These models are effective in controlling for individual factors that may lead to confounding, and adding year fixed effects helps to remove health and health plan related confounders. However, none of these models take into account the fact that gaining a discount and losing a discount may

¡ 0.001

Table 4.6: Effect of copay assistance status on pick-up rates per patient-price-year combination

	Model 1	Model 2	Model 3	Model 4
Copay	0.014***	0.014***	0.014***	0.011***
	(0.001)	(0.002)	(0.002)	(0.003)
Copay x Discount		0.00001		
		(0.00001)		
Copay x Final price			−0.00000	
			(0.00002)	
Encounter number				0.001***
				(0.0003)
Copay x Encounter number				0.001*
				(0.0004)
Patient-price-year fixed effects	Yes	Yes	Yes	Yes
Observations	144,049	144,049	144,049	144,049
Adjusted R^2	0.135	0.135	0.135	0.135

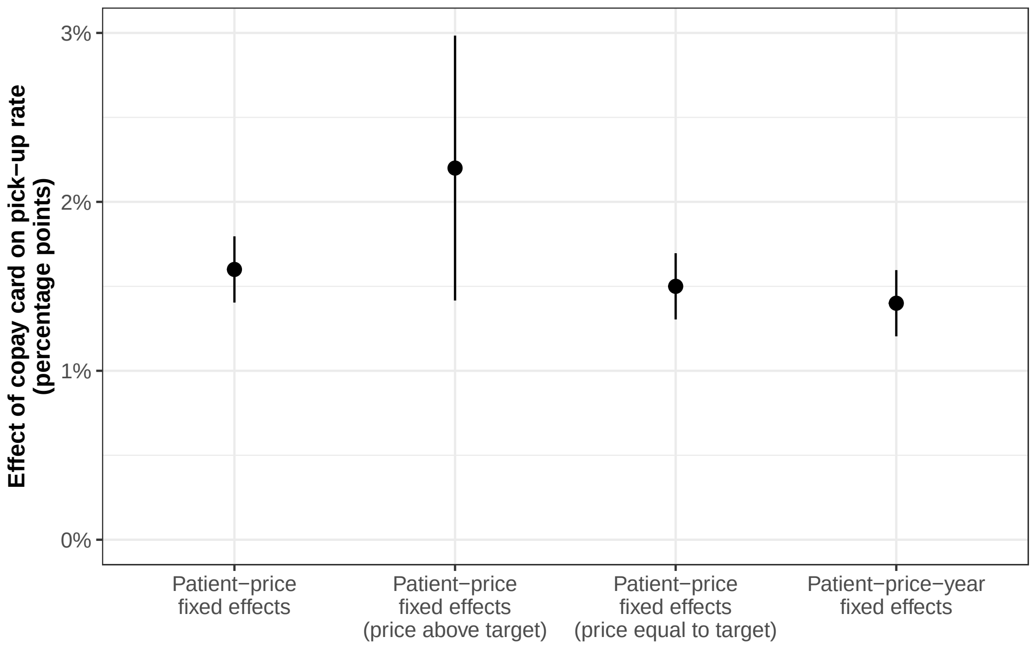
Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance status. The regression specifications include patient-price-year fixed effects. Definitions are as in Table 4.3.

be treated differently by decision makers. In this section, we examine whether the copay assistance effect differs between a patient who gains a discount and a patient who loses a discount.

In our data set, some patients start out receiving copay assistance, then lose their assistance, but pay the same out-of-pocket price because losing their discount co-occurs with a price drop. These patients lose a discount. Others see the reverse sequence: they do not receive assistance initially, then begin receiving assistance, but pay the same out-of-pocket price because the introduction co-occurs with a price hike. These patients gain a discount. Still others see multiple changes in copay assistance status (i.e., they gain and lose their discount at least one time each), but pay a constant out-of-pocket price throughout these changes in copay assistance status.

There are multiple psychological considerations relevant to how people will respond to gaining versus losing a discount. One consideration arises from the fact that there is a well-known asymmetry across gains and losses (Kahneman and Tverky, 1979): losses of a

Figure 4.2: Estimates of copay assistance effect on pick-up rates across specifications



Notes: The figure shows the effect of having copay assistance on pick-up rates (in percentage points) across our four main specifications: patient-price fixed effects, patient-price fixed effects where the price is above the target price, patient-price fixed effects where the price is equal to the target price, and patient-price-year fixed effects. Error bars indicate a 95 percent confidence interval around the estimated effect of copay assistance on pick-up rates.

given amount are more bad than gains of the same amount are good. If losing a discount destroys more transaction utility than getting a discount creates, we should see a bigger drop in demand among patients who lose a discount than we see a boost in demand among patients who gain a discount. Further, this reference dependence account would predict that experience with the drug (i.e., learning its therapeutic efficacy) should not impact the effect of changing copay assistance status.

An alternative to the reference dependence account arises from the fact that people often infer quality from price (Gneezy et al., 2014). Inferring quality from price means that, all else equal, people believe a more expensive version of a good is higher quality (Shiv et al., 2005). To the extent that people infer quality from price when evaluating medications, such a heuristic may also result in greater copay assistance effects among those who lose their discounts. To see why, consider that those who lose a discount first encounter a drug that is more expensive (even though they pay a discounted price). In these initial discounted pick-ups, patients believe they are getting a better drug than they otherwise would, given the price they are paying. When discount-losing patients actually lose their discounts, they lose: 1) the transaction utility provided by the discount and 2) the belief that they are getting a high(er) quality drug. On the other hand, discount-gaining patients initially believe they are getting a higher quality drug than they otherwise would, but when they gain the transaction utility afforded by the copay assistance, their assessment of drug quality also falls. Thus, where discount-losing patients experience additive negative effects of losing their copay assistance status, discount-gaining patients experience countervailing effects of gaining their copay assistance status.

For the above-described asymmetry in gaining/losing discounts to be true, we must assume patients believe pharmaceutical companies sometimes overcharge for their medications, but never undercharge. Such a (not unreasonable) assumption creates an asymmetry in judged medication quality across gained and lost discounts: on this assumption, patients would revise their quality assessments down when pre-discount prices fall, but do not revise them up when pre-discount prices rise. On this account, both the sequence of copay assistance status changes and experience with the drug should matter; in particular, we should see that as patients gain experience with a drug, the ordering of copay assistance status changes should matter less.

Table 4.7 shows the results. As expected, we find a much larger copay assistance effect when patients lose a discount (3.1 percentage points, $p < 0.001$) than when they gain a discount (0.1 percentage points, $p = 0.641$), for whom we do not find evidence of a copay assistance effect. Patients who experience multiple changes to their copay assistance status showed a significant, but intermediate, copay assistance effect (1.4 percentage points, $p < 0.001$). When we restrict our analysis to patients who see their first change in assistance status after at least three, or four, or five encounters, our results do not change materially. This suggests that people are not learning anything about the quality of the drug that is relevant to their purchasing decision.

4.5.5 Discussion

Our analysis of repeated purchases relies on discovering patients whose copay assistance status changes with no concomitant change in out-of-pocket costs. For example, someone

Table 4.7: Effect of copay assistance status on pick-up rates for different orderings

	Model 1	Model 2	Model 3
Copay assistance	0.031***	0.001	0.014***
	(0.002)	(0.002)	(0.002)
Order	Lose discount	Gain discount	More than 1 switch
Patient-price fixed effects	Yes	Yes	Yes
Observations	59,414	54,957	83,987
Adjusted R^2	0.132	0.150	0.131

Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance status for different orderings of copay (un)availability. Definitions are as in Table 4.3.

might apply a \$10 copay assistance discount to a drug whose list price is \$35, then lose that copay concurrently with a \$10 reduction in the listed price of the drug, such that the effective out-of-pocket cost remains \$25, despite losing their discount.

In a minimalist model examining that particular within-subject treatment, and as predicted by a transaction utility account of consumer decision making, we estimate a 1.6 percentage point increase in patients' propensity to pick-up their drugs when they receive copay assistance. This discount effect is robust to including year fixed effects and subsetting patients based on target price versus actual price. These robustness checks provide further evidence in favor of the transaction utility account: 1) the year fixed effects models handle health shocks and other time-varying confounds, and 2) the models in which we subset patients based on target versus actual price provide a test of alternative mechanisms to transaction utility, like salience or surprise accounts on which predictable or surprising prices cause increased pick-up rates, for which we find no evidence.

We do, however, uncover some nuance in how the sequencing of discounts matters: losing a discount reduces demand more than acquiring a discount increases it. In conjunction with our findings from the previous section on first-time purchase decisions, where we find that copay assistance decreases abandonment, our results suggests that providing patients with enduring discounts is likely to both increase the probability that their start their prescription regimen, and increase the chance that they stick to it. Thus, this section has provided evidence from several angles that discounts shift the demand curve for life-saving medications, and that this shift is best explained by discounts offering transaction utility. Nonetheless, we make these causal claims on the basis of estimates among a potentially unusual subset of the population: those who continuously pay one price for their prescription, but whose copay assistance status changes over time. While we have no reason to believe that this subset of people is psychologically special (and in some cases we know they are not; see our analysis of gained versus lost discounts), that possibility remains open, and we acknowledge the potential for objections to generalizing our estimates on the grounds of this external validity concern.

4.6 General Discussion

The impact of discounts on real-world consumer choice has been very hard to study, and for good reason. For one, pricing regulations in many industries make it difficult to find instances in which the discounted price of a good for one set of consumers is the same as the undiscounted price of that good for another set of consumers. Put differently, identifying shifts of the demand curve that result from discounts has been a challenge because such shifts are perfectly confounded with movements along the demand curve. Further, even if there were instances in which discounted and undiscounted prices were identical, it is almost always the case that only actualized purchasing decisions are observed; observational consumer data sets rarely, if ever, include both actualized and foregone purchases. Our data set solves both of these problems. The unique flexibility of pricing in the pharmaceutical market enables us to find instances in which discounted and undiscounted prices are identical—and this is sometimes true within an individual. And the unique nature of prescriptions means that we know when a patient is considering a purchase and when they choose not to make it.

Across two sets of analyses using this unique data set, we demonstrate that discounts, in the form of copay assistance, have a substantial positive effect on prescription pick-up rates. When we examine patients who were deciding whether to pick up their prescription for the first time, we estimate a 4.7 percentage point increase in pick-up rate in the presence of a discount. Similarly, when we examine repeated decisions by the same individuals, we observe a 1.6 percentage point increase in pick-up rates when copay assistance is available. These effect sizes are even more dramatic when one considers that 11 percent of prescription pick-ups are foregone.

Throughout, we have considered two competing models of a patient's decision to pick up their prescribed medication: the classical model, in which discounts merely move people along the demand curve, and the transaction utility model, in which discounts shift the demand curve itself. Those who have wanted to push back against the transaction utility model of consumer choice have had compelling arguments. Empirical work on transaction utility has traditionally been lab-based, so even when the phenomenon is identified, the purchasing scenario still is not naturalistic, decisions to buy are often either hypothetical or made with an endowment, and the goods are usually the target of discretionary spending. People do, in fact, have less incentive to act rationally when they are deciding in a non-naturalistic setting, where they are making hypothetical choices about their discretionary spending. This description stands in stark contrast, however, with our *in vivo* setting, in which people are spending their own money (or not) on drugs that might save their life (or not). If there were ever a context in which we would expect to observe rational consumer behavior, it would be this one.

4.A Appendix

This appendix contains two tables with auxiliary analyses. Below in Table 4.8, we show a number of variations in matching estimator constraints. For example, a patient picking up their prescription on Saturday, June 1, 2019 could be matched to a nearest neighbor who picked up their prescription six days earlier on Sunday, May 26, 2019 (exact week), or one who picked their prescription one day later on Sunday, June 2, 2019 (nearest week). The same is true for month of pick-up and age matching, as well as all combinations therein. Table 4.8 shows that while our estimates move around in response to these changes in the matching constraints, the qualitative results is preserved, and in most cases is quantitatively similar.

In Table 4.9, we show the effect of discount size on pick-up rates. To do so, we use our patient-price fixed effects approach to estimate the causal effect of an increase in the size of the discount received. In this analysis, all patients who made at least two pick-ups with copay assistance are included (hence the large number of observations, although many of those observations are filtered out by the fixed effects estimator because initial price, and therefore discount size does not vary); thus, this model estimates the effect of discount size on pick-up rates holding copay assistance status (and patient and final out-of-pocket cost) constant. We find a small, but statistically significant dose-dependent effect of discount size, such that pick-ups rates increase by 0.2 percentage points for every \$100 dollar increase in discount size. We take this as evidence that the presence of a discount matters more than its size. In other words, in the inequality we have been exploring in this chapter, $v(p) + d > v(c)$, d may be closer to a binary variable than a continuous one.

Table 4.8: Matching variations for the abandonment analysis

Estimate	SE	Effective obs.	Time-match	Age-match
0.06***	0.001	201	Exact month	Exact age
0.04***	0.001	4,397	Exact month	Nearest age
0.10***	0.001	43	Exact week	Exact age
0.04***	0.001	1,260	Exact week	Nearest age
0.04***	0.000	4,511	Nearest month	Exact age
0.05***	0.001	32,466	Nearest month	Nearest age
0.04***	0.000	4,511	Nearest week	Exact age
0.05***	0.001	32,466	Nearest week	Nearest age
0.04***	0.000	4,511	No date	Exact age
0.05***	0.001	32,466	No date	Nearest age

Notes: The table shows additional matching specifications for the abandonment analysis. The outcome variable is the likelihood that a patient purchases their newly prescribed medicine and the treatment variable is a patient's copay status, which takes the value of one if the patient receives copay assistance. *Time-match* specifies whether treatment and control patients are matched by the week or month of their purchase, or not by date at all. *Age-match* specifies whether patients are matched to someone with the exact same age or to the nearest-neighbor. All definitions are as in Table 4.2.

Table 4.9: Effect of copay assistance discount on pick-up rates per patient-price combination

Copay discount	0.00002***
	(0.00001)
Patient-price fixed effects	Yes
Observations	625,860
Adjusted R^2	0.116

Notes: The table reports the effects resulting from fixed effects linear probability models of purchase likelihood on copay assistance discount. *Copay discount* is the amount of discount provided by the copay assistance. The analysis is based only on purchases for which copay assistance is available. Definitions are as in Table 4.3.

Chapter 5

Incentives, Performance and Choking in Darts[1]

5.1 Introduction

Incentives are at the core of economics. From labor supply to crime, and from consumption to education, incentives play a central role in economic theory. An important prediction is that people exert more effort when they face stronger incentives, and that this increased effort in turn leads to better performance. Various experimental studies have shown that higher monetary incentives indeed improve performance (Smith and Walker, 1993; Camerer and Hogarth, 1999; Gneezy and Rustichini, 2000). Outside the behavioral laboratory, research has similarly shown that workers produce more output when they are paid a piece rate rather than an hourly rate (Paarsch and Shearer, 1999; Prendergast, 1999; Lazear, 2000; Shearer, 2004) and that students perform better at tests when they are paid according to their performance (Levitt et al., 2016).

An alternative line of research, predominantly in psychology, suggests that higher incentives do not always improve performance and can even backfire (Kamenica, 2012). High incentives can cause people to consciously think about their actions in otherwise automatically performed tasks, and consequently impede performance (Baumeister, 1984;

[1]Published as 'Klein Teeselink, B., R. J. D. Potter van Loon, M. J. van den Assem, and D. van Dolder (2020). Incentives, Performance and Choking in Darts. *Journal of Economic Behavior and Organization 57(1)*, 115-129'.

Beilock and Carr, 2001; Masters and Maxwell, 2008). Also, higher incentives can lead to more arousal. The Yerkes-Dodson law postulates that arousal has a non-monotonic effect on performance: moderate increases enhance performance, whereas the effect turns negative if arousal surpasses a critical threshold that depends on task difficulty (Yerkes and Dodson, 1908). Ariely et al. (2009) show that high stakes harm performance in a diverse set of experimental tasks that draw on subjects' memory, creativity, or motor skills, but not in a task that solely requires physical effort. Inferior performance in the presence of incentives for superior performance is known as choking under pressure (Baumeister and Showers, 1986; Beilock, 2010).

The present chapter contributes to the understanding of the link between incentives and performance by analyzing four large data sets of competitive darts matches of professional, amateur and youth players. Darts offers a unique combination of attractive features for this type of research. The game is played in a real-world natural environment. Like many other real-life activities—such as the work of pilots, bus drivers, soldiers, surgeons and dentists—the task is neither entirely physical nor entirely mental, but combines elements of both. The sustained concentration demanded by the game can be regarded as a form of effort. Just as with controlled experiments, the game has a clearly defined set of rules and objectives. Performance can be observed at the individual level, and unlike many other field tasks it is not confounded by behavior of others, such as colleagues or competitors. As opposed to many other field settings—where it can be rational for risk-averse individuals to shift to low-risk low-reward strategies when the stakes increase—the optimal approach to darts is independent of risk preferences, because a rational player, no matter how risk averse, will always try to maximize her probability of winning.

By using darts data, we connect to a broader literature that uses sports data to investigate economic hypotheses. Examples are the study of discrimination in basketball (Price and Wolfers, 2010), principal-agent theory in cricket (Gauriot and Page, 2015), and mixed-strategy Nash equilibria in soccer penalty shootouts (Chiappori et al., 2002). The prediction that athletes perform better under higher monetary incentives has been confirmed for various sports, including golf (Ehrenberg and Bognanno, 1990*a*,*b*), auto racing (Becker and Huselid, 1992), horse racing (Lynch, 2005), and tennis (Gilsdorf and Sukhatme, 2008*a*,*b*). Choking under pressure has previously been observed in, for example, penalty shootouts in soccer (Dohmen, 2008), free throws in basketball (Cao et al., 2011; Goldman and Rao, 2012; Toma, 2017; Böheim et al., 2019), putting in golf (Hickman and Metz, 2015), shooting in biathlon (Lindner, 2017; Harb-Wu and Krumer, 2019), and tennis (Paserman, 2010; Cohen-Zada et al., 2017).

To investigate the effect of incentives on performance in darts, we exploit naturally occurring within-match variation in the benefit (cost) of throwing well (poorly). We find that amateur and youth players display a sizable performance decrease at decisive moments. Professional players appear less susceptible of such choking under pressure.

In the remainder of the chapter we explain the game of darts and our data (Section 5.2), present the analyses and results (Section 5.3), and conclude (Section 5.4).

5.2 Description Darts and Data

5.2.1 Darts

In darts, two players compete with each other by sequentially throwing darts at a dartboard. A dartboard is divided into areas that represent points in the range of 1 to 20 (see Figure 5.1). The number of points is doubled when a dart is thrown in the outer band, and tripled when it is thrown in the inner band. The outer ring in the center of the dartboard (or 'outer bull') gives 25 points, the inner circle (or 'inner bull') gives 50 points. The maximum score with one dart is 60 points (triple 20).

Darts matches are played in either 'leg' or 'set' format. A match in leg format is a best-of-n contest, where each of the n sub-contests is called a leg. A match in set format also is a best-of-n contest, but each sub-contest is then called a set, which in turn is a best-of-n contest with legs. A set thus resembles a match in leg format.

Players normally start a leg with 501 points each, and take turns to throw three darts. One turn of three darts is commonly referred to as one 'throw'. The sum of the points in a throw is subtracted from the remaining number of points. To finish and win a leg, a player is required to reach zero (exactly) by hitting either a 'double' or the inner bull. For example, a player with 18 points remaining can finish by hitting double 9. If the score of a dart exceeds the number of points the player has left, her entire throw of three darts is

Figure 5.1: Dartboard

rendered invalid. Players take turns to start legs. The starter of the first leg is generally determined by shots at the center of the dartboard, with the player closest starting.

5.2.2 Data

Our data are from Darts for Windows (www.dartsforwindows.com). Darts for Windows collects data from various sources, most notably from darts associations and darts competitions that use the Darts for Windows computer software. There are four categories of data: Youth, Super League, British Inter-County Championship (BICC), and International. We downloaded all available International data on July 11, 2017, all available BICC data on November 26, 2017, and all available data for the other two categories on July 4, 2019.

The four data sets cover the matches of different types of players, ranging from amateur youth players to professional adults. The international tournaments sample covers 15,205 matches between professional players from 1974-2017, and includes matches played at famous tournaments such as the UK Open and the PDC World Darts Championship. The BICC is a competition between amateur players from various counties in the United Kingdom. The BICC sample comprises 10,369 matches played in the period 2005-2017.

The Super League is a regional amateur league that is played mostly in the United Kingdom. The Super League sample contains 1,643 matches from 2007-2019. The Youth sample consists of 2,164 matches from tournaments for boys under 18, boys under 21, and girls under 21, that took place in the period 2001-2019.

For each match, we have granular data down to each player's score in one throw of three darts. Along with the score per throw, we know the date of the match, players' names, and the starter of the first leg.

In our analyses we treat sets as separate matches. Most matches in the Youth (51 percent), Super League (93 percent), and BICC (100 percent) sample are played in set format. These matches are virtually always between two teams, where each set is played by a different team member. Such sets can therefore rightfully be regarded as matches on their own. In the International sample, only a small proportion (5.6 percent) of the matches are in set format, and sets are generally played by the same player. For consistency, we nevertheless similarly treat these sets as separate matches. To make sure that our results are not sensitive to this approach, we also conduct robustness analyses that exclude the data from matches in the International sample that were played in set format. Treating sets as matches increases the total number of matches in the four data sets combined from 29,381 to 126,440.

We exclude matches with legs that do not start at 501 points, matches where one or more scores are missing, matches with more than two players, and matches where both players have the same name. After these cleaning operations, 125,679 matches remain.

Table 5.1 presents summary statistics. In total, our data comprise more than half a million legs and more than eight million throws. There are clear skill differences between the four categories, with average points per throw ranging from 49.0 (Youth) to 65.8 (International), and the percentage of successful one-dart finishes ranging from 31.4 (Youth) to 46.5 (International).[2] The skill differences are also reflected in the average leg length and in the proportion of starting players winning the leg, with better players taking fewer throws and being more likely to benefit from throwing first. The proportion of successful one-dart finishes is higher if we consider only the first opportunity of every leg instead

[2]A one-dart finish opportunity is a throw where a player can finish the leg with just one dart. The remaining number of points in such a situation is 2, 4, 6, ..., 36, 38, 40 or 50.

Table 5.1: Summary statistics

	International	BICC	Super League	Youth
Matches	17,855	89,019	12,195	6,610
Legs	85,446	382,195	51,071	20,054
Throws	1,200,383	5,922,129	816,745	378,481
Players	2,644	10,788	5,414	2,978
Legs per match	4.79	4.29	4.19	3.03
Throws per leg	14.0	15.5	16.0	18.9
Points per throw (all)	65.8	60.1	57.7	49.0
Points per throw (first three)	79.6	70.6	67.5	59.3
Starter wins leg (proportion)	0.587	0.570	0.560	0.548
One-dart finish opportunities	111,241	587,876	81,385	45,893
One-dart finish opportunities (first only)	68,547	345,416	46,518	20,563
Successful one-dart finish (proportion)	0.465	0.425	0.419	0.314
Successful one-dart finish (proportion, first only)	0.528	0.473	0.465	0.363

Notes: Matches, *Legs*, *Throws* and *Players* are the number of matches, legs, throws, and players, respectively. *Legs per match* is the average number of legs per match. *Throws per leg* is the average number of throws by both players combined per leg. *Points per throw (all)* is the average number of points per throw across all throws. *Points per throw (first three)* is the average number of points per throw across players' first three throws in every leg. *Starter wins leg* is the proportion of legs won by the player who started the leg. *One-dart finish opportunities (first only)* is the number of throws where a player can finish the leg with one dart (for the first time in that leg). *Successful one-dart finish (first only)* is the proportion of throws where a player could finish the leg with one dart (for the first time in that leg) and finished the leg in the given throw.

of all. This difference can be attributed to a selection effect, because better players are more likely to be successful on their first attempt.

5.3 Analyses and Results

The incentive to do well in darts varies both across and within matches. Across matches, players will be motivated by the amount of prestige, prize money, and media attention. Compare, for example, the final of the internationally televised PDC World Darts Championship where the 2019 winner (Dutchman Michael van Gerwen) took home £500,000, with a match in the first round of the men's singles tournament of the Lincolnshire Family Darts Festival where the 2019 winner cashed £2,000. Our data, however, does not contain sufficient information to systematically proxy for such variation.

Within matches, there is considerable variation in the impact of the quality of a dart throw on the likelihood of winning the match, and players can be expected to adjust their effort provision accordingly (Konrad, 2009). The incentive to throw well is relatively high when both players are close to winning a leg, and when both players are close to winning the match. Our analyses exploit this naturally-occurring within-match variation, and consider its effect on players' finishing performance (Section 5.3.1) and on the points they throw in the first three throws of a leg (Section 5.3.2).

5.3.1 Finishing

We first examine how players' finishing performance is affected when both players can win the match by winning the current leg. Such legs are highly consequential because poor performance can irreversibly result in losing the match, whereas strong performance can secure the win.[3] Second, we examine how players' finishing performance is affected when their opponent can finish the leg in the subsequent throw. The pressure on a player to finish is relatively high when her opponent can also finish, and relatively low otherwise. Last, we consider the interaction of these conditions. In such critical cases, where both can win the match by winning the current leg and where the opponent can finish in the subsequent throw, the pressure to do well is especially high.

We exclusively analyze situations where the player can finish the leg with one dart. This is the case when she has 2, 4, 6, ..., 36, 38, 40 or 50 points left. Strategy plays no role in these situations, because any approach other than trying to finish the leg in the current throw is sub-optimal. In contrast, if a player needs multiple darts to finish she might instead try to maneuver herself into a better finishing position for the next throw. Such a strategy can be attractive in situations where the opponent is unlikely or unable to finish, and would thus generate the false impression of lower performance in lower-incentive situations.

We use a fixed effects logit approach to regress finishing performance on incentives. We control for skill differences between players and across matches through player-match fixed effects, and for possible warming-up and fatigue effects within a match by including a polynomial of order n for the player's number of throws in the match prior to the current throw, where the value of n is chosen to minimize the AIC.[4] When a player displays

[3]We intentionally only consider situations where *both* can win the match, and not those where only one can win. First, if only one can win, the incentive to do well can be both high and low, depending on the closeness of the match. Second, using situations where only one can win would lead to biased coefficient estimates due to regression to the mean: players who are ahead (behind) have on average been (un)lucky with their previous throws, and as a consequence their performance will in expectation decrease (increase).

[4]If the match has a set format, the number of throws includes the throws of the player in previous sets.

Table 5.2: Regression results for the likelihood of finishing

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel A: International						
Decisive leg	-0.025^{***}			-0.006	0.026	0.026
	(0.008)			(0.008)	(0.026)	(0.026)
Opp. can finish		0.002			0.003	
		(0.005)			(0.005)	
Opp. can finish with 1 dart			-0.005			-0.003
			(0.005)			(0.006)
Opp. can finish with 2 darts			0.010^{*}			0.011^{*}
			(0.005)			(0.006)
Opp. can finish with 3 darts			-0.001			-0.0004
			(0.006)			(0.006)
Decisive leg x Opp. can finish					-0.036	
					(0.028)	
Decisive leg x Opp. can finish with 1 dart						-0.048
						(0.029)
Decisive leg x Opp. can finish with 2 darts						-0.033
						(0.030)
Decisive leg x Opp. can finish with 3 darts						-0.008
						(0.034)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	3	3	3	3	3	3
McFadden pseudo R-squared	0.139	0.085	0.085	0.085	0.085	0.085
Observations	111,241	68,547	68,547	68,547	68,547	68,547
Effective observations	84,994	37,837	37,837	37,837	37,837	37,837

Notes: The table reports the average marginal effects resulting from logit regression analyses of finishing performance across throws where the player can finish the leg with one dart. Model 1 uses all one-dart finish opportunities, whereas Models 2-6 use players' first one-dart finish opportunity in a leg only. The dependent variable takes the value of 1 if the player finishes in the given throw, and 0 otherwise. *Decisive leg* is a dummy variable that takes the value of 1 if both players can win the match by winning the current leg. *Opp. can finish* (*with 1, 2, or 3 darts*) is a dummy variable that takes the value of 1 if the player's opponent can finish the leg in the subsequent throw (and needs one, two, or three darts). The regression specifications include player-match fixed effects, and a polynomial of order n for the player's prior number of throws in the match, where n is chosen to minimize the AIC. Average marginal effects are corrected for incidental parameter bias (Fernández-Val, 2009). *Effective observations* is the number of observations that contribute to the estimation of the underlying regression coefficients, which can be found in Table 5.5 in the Appendix. Standard errors are in parentheses. Asterisks denote significance at the 0.01 (***), 0.05 (**) and 0.1 (*) level.

no variation in one-dart finishing success within a match, her observations effectively do not contribute to the estimations of the coefficients; the average marginal effects that we present, however, are based on all observations.[5]

Table 5.2 presents the regression results.[6] Model 1 shows that the effect of *Decisive leg* is negative and significant at the one-percent level in all samples. A player's finishing probability on average deteriorates by 2.5-7.7 percentage points if both players can win the match by winning the leg, compared to situations where none or only one of them is close to winning the match. Scaled by the sample-specific proportion of successful

[5]In doing so we follow the conventional approach. Calculating average marginal effects on the basis of only the effective observations would amplify the effect sizes.

[6]Table 5.5 in the Appendix displays the underlying coefficient estimates. The p-values of the coefficients are higher than those of the average marginal effects, but support the same general conclusions.

Table 5.2: Continued

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel B: BICC						
Decisive leg	-0.077^{***}			-0.059^{***}	-0.007	-0.008
	(0.003)			(0.003)	(0.008)	(0.008)
Opp. can finish		0.003			0.007^{***}	
		(0.002)			(0.003)	
Opp. can finish with 1 dart			0.001			0.007^{***}
			(0.003)			(0.003)
Opp. can finish with 2 darts			0.003			0.006^{**}
			(0.003)			(0.003)
Opp. can finish with 3 darts			0.006^{**}			0.008^{***}
			(0.003)			(0.003)
Decisive leg x Opp. can finish					-0.058^{***}	
					(0.008)	
Decisive leg x Opp. can finish with 1 dart						-0.079^{***}
						(0.009)
Decisive leg x Opp. can finish with 2 darts						-0.042^{***}
						(0.009)
Decisive leg x Opp. can finish with 3 darts						-0.032^{***}
						(0.010)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	4	3	3	4	4	4
McFadden pseudo R-squared	0.124	0.072	0.072	0.073	0.073	0.073
Observations	587,876	345,416	345,416	345,416	345,416	345,416
Effective observations	463,690	193,135	193,135	193,135	193,135	193,135
Panel C: Super League						
Decisive leg	-0.035^{***}			-0.035^{***}	-0.001	-0.001
	(0.009)			(0.009)	(0.023)	(0.023)
Opp. can finish		0.003			0.005	
		(0.006)			(0.006)	
Opp. can finish with 1 dart			0.004			0.007
			(0.007)			(0.007)
Opp. can finish with 2 darts			-0.003			-0.002
			(0.007)			(0.007)
Opp. can finish with 3 darts			0.009			0.013^{*}
			(0.007)			(0.008)
Decisive leg x Opp. can finish					-0.039	
					(0.024)	
Decisive leg x Opp. can finish with 1 dart						-0.052^{**}
						(0.026)
Decisive leg x Opp. can finish with 2 darts						-0.013
						(0.027)
Decisive leg x Opp. can finish with 3 darts						-0.055^{*}
						(0.030)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	4	3	3	3	3	3
McFadden pseudo R-squared	0.126	0.074	0.074	0.074	0.074	0.075
Observations	81,385	46,518	46,518	46,518	46,518	46,518
Effective observations	65,063	26,536	26,536	26,536	26,536	26,536

one-dart finishes (see Table 5.1), the negative impact on performance ranges between 5.3 percent for International players and 18.1 percent for BICC players. For Super League and Youth the performance decreases are 8.4 and 15.3 percent, respectively.

Table 5.2: Continued

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel D: Youth						
Decisive leg	-0.048^{***}			-0.028^{***}	0.007	0.007
	(0.010)			(0.010)	(0.026)	(0.026)
Opp. can finish		-0.032^{***}			-0.028^{***}	
		(0.008)			(0.009)	
Opp. can finish with 1 dart			-0.045^{***}			-0.039^{***}
			(0.009)			(0.010)
Opp. can finish with 2 darts			-0.030^{***}			-0.030^{***}
			(0.009)			(0.010)
Opp. can finish with 3 darts			-0.020^{**}			-0.020^{*}
			(0.010)			(0.010)
Decisive leg x Opp. can finish					-0.039	
					(0.026)	
Decisive leg x Opp. can finish with 1 dart						-0.065^{**}
						(0.028)
Decisive leg x Opp. can finish with 2 darts						-0.016
						(0.029)
Decisive leg x Opp. can finish with 3 darts						-0.010
						(0.033)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	3	4	4	4	4	3
McFadden pseudo R-squared	0.133	0.061	0.061	0.061	0.061	0.062
Observations	45,893	20,563	20,563	20,563	20,563	20,563
Effective observations	36,385	9,322	9,322	9,322	9,322	9,322

To analyze how finishing performance is affected by the opponent's opportunity to finish the leg in the subsequent throw, we need to restrict the data to players' first finishing attempt in every leg. The inclusion of subsequent attempts would lead to biased estimates, because the random component in a player's performance in a match affects both the estimated "normal" skill of the player (as captured through the player-match fixed effects) and the likelihood that a situation arises where the opponent can finish. For example, if a player misses a finish opportunity, her finishing statistic for the match worsens, and if she then gets a new opportunity to finish, her opponent will be closer to finishing because of the extra turn. Hence, including subsequent attempts would positively bias the estimated effect of the opponent's opportunity to finish on performance.

Model 2 shows how a player's finishing probability changes if her opponent can finish the leg in the next throw, compared to situations where her opponent cannot finish. The pressure from an opponent who can finish deteriorates performance among Youth players only. Youth players are 3.2 percentage points less likely to finish as compared to situations where their opponent cannot finish. In the other three samples, the average marginal effects are close to zero and statistically insignificant.

Situations where the opponent can finish consist of three rather different types of cases, where finishing requires either one, two or three darts. Finishing difficulty increases with this minimum number of darts that are required. Three-dart finishes are particularly difficult, and an opponent in such a position consequently poses relatively little threat. Model 3 therefore distinguishes between situations where the opponent needs one, two or three darts. With this more fine-grained approach, the effect of an opponent who can finish generally remains statistically insignificant for professional and amateur players, even if the opponent's finishing difficulty is relatively low because she can finish with a single dart.[7] Among Youth players the performance deterioration increases with the pressure. Compared to situations where the opponent cannot finish, Youth players are 2.0, 3.0, and 4.5 percentage points less likely to finish when their opponent needs three darts, two darts, or one dart, respectively.

For completeness, Model 4 re-estimates the effect of *Decisive leg* for the restricted samples. The average marginal effects now range between -0.6 and -5.9 percentage points, and are consistently smaller than those of Model 1. This difference can be attributed to selection effects: the omitted data most likely contained a disproportionate amount of observations from players who choked under the pressure of a decisive leg, and from situations with additional pressure because the opponent was closer to finishing.

Model 5 combines the pressure effects of decisive legs and opponents who can finish, and also includes the interaction of the two. The pressure on a player is presumably highest when the two conditions apply simultaneously. Table 5.3 presents their joint effect on performance, which is equal to the sum of the marginal effects of *Decisive leg*, *Opp. can finish* and their interaction. Amateur and youth players display sizeable choking effects: BICC, Super League and Youth players perform 5.8, 3.5 and 6.0 percentage points worse, respectively, in situations where both conditions apply than in situations where neither apply. The three differences are statistically significant (all $p < 0.001$). Scaled by the sample-specific proportion of one-dart finishes (see Table 1), the effect sizes translate into performance deteriorations of 12.3, 7.5 and 16.5 percent, respectively.

[7] Ötting et al. (2020) analyze a data set that exclusively consists of about one year of professional darts matches organized by the Professional Darts Cooperation. For situations where a player can finish with one dart, they similarly find no evidence that performance is affected by her opponent's finish opportunity.

Table 5.3: Joint effects of decisive leg and opponent close to finishing

	Marginal effect	z-value	p-value
Panel A: International			
1/2/3 darts	-0.007	-0.657	0.511
1 dart	-0.025	-1.813	0.070
2 darts	0.004	0.280	0.779
3 darts	0.018	0.864	0.387
Panel B: BICC			
1/2/3 darts	-0.058	-14.372	0.000
1 dart	-0.080	-16.279	0.000
2 darts	-0.044	-8.293	0.000
3 darts	-0.032	-4.599	0.000
Panel C: Super League			
1/2/3 darts	-0.035	-3.188	0.001
1 dart	-0.046	-3.236	0.001
2 darts	-0.017	-1.130	0.259
3 darts	-0.044	-2.215	0.027
Panel D: Youth			
1/2/3 darts	-0.060	-4.750	0.000
1 dart	-0.094	-6.130	0.000
2 darts	-0.039	-2.271	0.023
3 darts	-0.023	-1.034	0.301

Notes: The table reports the sum of the average marginal effects of *Decisive leg*, *Opp. can finish* (*with 1, 2, or 3 darts*) and the interaction of these two variables, according to Models 5 and 6 in Table 5.2. Underlying coefficients can be found in Table 5.6 in the Appendix.

Model 6 similarly combines the two types of pressure effects, but distinguishes between situations where the opponent needs one, two, or three darts. Table 5.3 presents the joint effect sizes, which correspond to the difference in performance between situations where the leg is decisive and the opponent needs one, two or three darts, and situations where the leg is not decisive and the opponent cannot finish in the next throw. In line with the previous results, BICC, Super League, and Youth players display significant choking under pressure at the end of a close match. As expected, these choking effects are largest when the opponent can secure the win with only one dart. In this extreme situation, BICC, Super League and Youth players perform 8.0, 4.6 and 9.4 percentage points worse, respectively. The corresponding scaled deteriorations are 16.9, 9.9 and 25.8 percent.

In the category of International matches, with relatively skilled and experienced players, there is no compelling evidence of choking in decisive legs where opponents are close to finishing. Based on Model 5, the overall decrease in performance equals an insignificant 0.7 percentage points. According to Model 6, International players are 2.5 percentage points less likely to finish at the end of a close match when the opponent can secure the win with a single dart. Statistically, this difference is only marginally significant.

Note that when both types of incentive variables and their interaction are included, the marginal effects of *Decisive leg* are relatively close to zero as compared to those in Model 4. This is not surprising, because in Model 5 and Model 6 the marginal effects for this variable exclusively refer to situations where the player is significantly ahead in the leg: the player can finish with one dart, while the opponent cannot finish in the next throw. In such situations, players do not have a particularly strong incentive to perform well. Furthermore, the marginal effects of the opponent's finish opportunity in Model 5 and Model 6 are similar to those in Model 2 and Model 3, which makes sense because decisive legs are only a fraction of the total number of legs. Last, the interaction effects are always negative and sometimes statistically significant, which suggests that the adverse impact of a decisive leg and that of an opponent being close to finishing amplify each other.

One possible concern about the previous results is that variation in one-dart finishing difficulty is not accounted for. Compare, for example, a player with 2 and a player with 40 points remaining. Both can finish with one dart (by hitting double 1 and double 20, respectively). If the player with 2 points misses and instead throws 1 point, her throw is over. If the player with 40 points misses and instead throws 20 points, she still has the opportunity to finish with her second or third dart (e.g., by hitting double 10). We can control for finishing difficulty by expanding the regression models with fixed effects for the number of points that the player has left at the start of her throw. Table 5.7 in the Appendix shows that including points-left fixed effects does not materially affect the results.

Another possible concern relates to our treatment of set-format matches and the relatively weak evidence of choking for International matches. Throughout our analyses, we have treated sets as separate matches, which especially makes sense if each set is played by a different player of a team. In international tournaments, however, set-format matches are generally entirely played between the same two players, and winning a set is therefore substantially less important than winning the match. Treating sets as matches may consequently have diluted possible evidence of choking effects at truly decisive moments in this category of data. Table 5.8 in the Appendix shows the results for the International

Table 5.4: Regression results for the number of points thrown

	Model 1: Throw 1	Model 2: Throw 2	Model 3: Throw 3	Model 4: Throws 1-3
Panel A: International				
Decisive leg	−0.246	−0.992**	−0.432	−0.533*
	(0.485)	(0.486)	(0.484)	(0.279)
Player-match fixed effects	Yes	Yes	Yes	Yes
Throw fixed effects	-	-	-	Yes
Polynomial throw number (order)	9	5	9	7
Observations	170,892	170,892	170,884	512,668
Adjusted R^2	0.234	0.238	0.232	0.241
Panel B: BICC				
Decisive leg	−1.732***	−1.943***	−1.922***	−1.735***
	(0.205)	(0.206)	(0.206)	(0.118)
Player-match fixed effects	Yes	Yes	Yes	Yes
Throw fixed effects	-	-	-	Yes
Polynomial throw number (order)	9	9	9	9
Observations	764,390	764,390	764,384	2,293,164
Adjusted R^2	0.129	0.130	0.131	0.134
Panel C: Super League				
Decisive leg	−1.722***	−0.661	−0.347	−0.948***
	(0.544)	(0.542)	(0.556)	(0.315)
Player-match fixed effects	Yes	Yes	Yes	Yes
Throw fixed effects	-	-	-	Yes
Polynomial throw number (order)	5	4	9	5
Observations	102,142	102,142	102,132	306,416
Adjusted R^2	0.123	0.122	0.113	0.123
Panel D: Youth				
Decisive leg	−0.306	−1.035	−0.793	−0.791**
	(0.685)	(0.687)	(0.691)	(0.379)
Player-match fixed effects	Yes	Yes	Yes	Yes
Throw fixed effects	-	-	-	Yes
Polynomial throw number (order)	6	4	9	2
Observations	40,106	40,106	40,092	120,304
Adjusted R^2	0.139	0.147	0.144	0.151

Notes: The table reports the coefficients resulting from OLS regression analyses of the number of points thrown in the first, second, and/or third throw of a leg. The regression specifications include a polynomial of order n for the player's prior number of throws in the match, where n is chosen to maximize the adjusted R-squared. Model 4 includes throw fixed effects to control for differences in the average across throws. Other definitions are as in Table 5.2.

sample after excluding all set-format matches. Notwithstanding the sensitivity of some average marginal effects, all previous conclusions remain the same.

5.3.2 Points Thrown

We now turn to the early stage of legs, to examine how point-throwing performance is affected if both players can win the match by winning the current leg. As explained in the previous section, such legs are highly consequential because poor performance can irreversibly result in losing the match, whereas strong performance can secure the win. We consider each player's first three throws in every leg only.[8] Strategy plays no role in

[8] In exceptional cases where the opponent started the leg and won it in three throws, we only consider players' first two throws.

these first few throws, because any approach other than trying to throw as many points as possible is sub-optimal.[9] In contrast, in subsequent throws a player may try to maneuver herself into a favorable finishing position, for example by trying to reach a remaining score of 40.[10] As in the previous analysis, we control for skill differences between players and through time by including player-match fixed effects, and for possible warming-up and fatigue effects within a match by including a polynomial of order n for the player's number of throws in the match prior to the current throw, where the value of n is chosen to maximize the adjusted R-squared.

Table 5.4 presents the OLS regression results. The first three models show how the number of points thrown is affected if both players can win the match by winning the leg, for each of the first three throws separately. The coefficients are consistently negative, but not always statistically significant. Model 4 shows the results for the first three throws combined. For completeness, this model includes throw fixed effects to control for small differences in the average performance across throws. Again, there is clear evidence of choking under pressure. In each of the four samples, throwing performance worsens significantly when the leg is decisive. The effect size is largest for BICC, and smallest (and only marginally significant) for International players. Scaled by the average number of points per throw in the first three throws (see Table 5.1), the coefficients mean that players' performance deteriorates by 2.5 and 0.7 percent in these two samples. For Super League and Youth, the relative performance decreases are 1.4 and 1.3 percent, respectively.

Following our approach in the previous section, we have also conducted the analyses using International leg-format matches only, to alleviate the concern that set-format

[9] In theory, some slight strategic adjustments are conceivable. To maximize the expected number of points, a high-skilled player should aim at triple 20 and a lower-skilled player should aim at triple 19 (Tibshirani et al., 2011). However, when playing against a substantially lower-skilled player, a high-skilled player may want to aim at triple 19 to reduce variance and thereby reduce the small probability of losing due to a streak of bad luck. Similarly, when playing against a substantially higher-skilled player, a low-skilled player may want to aim at triple 20 to increase variance in the hope of being lucky. In practice, however, players tend to aim at triple 20 regardless of their skill. Note that such strategic considerations should be independent of risk preferences for any player who wants to maximize her winning probability, and, more importantly, that there is no reason for such strategic adjustments to vary systematically between decisive and non-decisive legs.

[10] Technically, players may want to aim for less than the maximum number of points in throw three already if they did exceptionally well in the first two throws. Such situations are, however, extremely rare. Furthermore, our strategy of using the first three throws and excluding all subsequent throws is empirically supported by the average number of points: this statistic is stable across throw one, two and three, and decreases from throw four onwards.

matches diluted the possible evidence of choking among professional players. Table 5.9 in the Appendix gives the results. For the first three throws combined, the choking effect is slightly more pronounced as compared to the original results, but the effect size remains relatively small.

5.4 Conclusions and Discussion

The present chapter examines how naturally-occurring within-match variation in the incentive to perform well impacts performance in darts. The game of darts offers an attractive naturally-occurring research setting, because performance can be observed at the individual level and without obscuring effects of risk considerations and behavior of others. Like many other real-life activities, playing darts is neither entirely physical nor entirely mental, but combines elements of both. We use four large data sets that cover the matches of different categories of players, ranging from amateur youth players to professional adults. We analyze how players perform in the early stage of a leg, when they have to throw as many points as possible, and in the final stage of a leg, when they have to finish by throwing a so-called ‘double’.

Among youth and amateur players, performance deteriorates substantially if a player and her opponent are close to winning the match. More specifically, if both can win the match with just one dart, a player is 5-9 percentage points less likely to finish as compared to situations that do not bring this additional pressure. Relative to the sample-specific averages, these effect sizes mean that performance worsens by 10-26 percent. Such choking under pressure also occurs at the start of decisive legs, albeit to a weaker extent. Scaled by the sample-specific averages, youth and amateur players throw 1.3-2.5 percent fewer points in their first three throws if both players can win the match by winning the leg.

Professional players appear less susceptible of choking under the high pressure of decisive legs than youth and amateur players: there is limited evidence of deteriorating finishing performance, and for the number of points thrown the adverse impact is relatively small. The different findings for this category suggest that choking under pressure can be mitigated by training and decreases with experience. At the same time, however, because the ability to deal with pressure is a competitive advantage, the low sensitivity of professional players can also be the result of a selection effect. Most of the earlier studies that have found evidence for choking under pressure in sports focused on professionals. Our analysis uses a broader set of players, and suggests that choking under pressure is (even) more of a concern for lower-skilled individuals.

Our results speak to a growing literature on the limits of increasing incentives as a recipe for better performance. Conflicting with the classical prediction in economics that higher incentives improve performance, darts players display clear symptoms of choking under pressure. Our conjecture is that the additional pressure leads players to consciously think about their actions, which disrupts the normal automatic processing of the well-trained task of throwing darts.

5.A Appendix

Table 5.5: Coefficient estimates underlying the average marginal effects in Table 5.2

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel A: International						
Decisive leg	-0.128^{**}			-0.030	0.138	0.137
	(0.053)			(0.070)	(0.219)	(0.219)
Opp. can finish		0.009			0.015	
		(0.041)			(0.041)	
Opp. can finish with 1 dart			-0.025			-0.015
			(0.045)			(0.045)
Opp. can finish with 2 darts			0.051			0.056
			(0.045)			(0.046)
Opp. can finish with 3 darts			-0.003			-0.002
			(0.049)			(0.050)
Decisive leg x Opp. can finish					-0.187	
					(0.231)	
Decisive leg x Opp. can finish with 1 dart						-0.251
						(0.244)
Decisive leg x Opp. can finish with 2 darts						-0.171
						(0.251)
Decisive leg x Opp. can finish with 3 darts						-0.041
						(0.278)
Panel B: BICC						
Decisive leg	-0.395^{***}			-0.291^{***}	-0.036	-0.039
	(0.020)			(0.027)	(0.068)	(0.068)
Opp. can finish		0.015			0.035^{*}	
		(0.019)			(0.020)	
Opp. can finish with 1 dart			0.007			0.036^{*}
			(0.021)			(0.021)
Opp. can finish with 2 darts			0.017			0.030
			(0.021)			(0.022)
Opp. can finish with 3 darts			0.029			0.039
			(0.023)			(0.024)
Decisive leg x Opp. can finish					-0.286^{***}	
					(0.070)	
Decisive leg x Opp. can finish with 1 dart						-0.393^{***}
						(0.074)
Decisive leg x Opp. can finish with 2 darts						-0.206^{***}
						(0.076)
Decisive leg x Opp. can finish with 3 darts						-0.158^{*}
						(0.085)

Notes: The table reports the coefficient estimates resulting from logit regression analyses of finishing performance across throws where the player can finish the leg with one dart. Model 1 uses all one-dart finish opportunities, whereas Models 2-6 use players' first one-dart finish opportunity in a leg only. Definitions are as in Table 5.2.

Table 5.5: Continued

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel C: Super League						
Decisive leg	-0.176^{***}			-0.172^{**}	-0.004	-0.007
	(0.052)			(0.071)	(0.185)	(0.185)
Opp. can finish		0.013			0.025	
		(0.048)			(0.049)	
Opp. can finish with 1 dart			0.020			0.035
			(0.053)			(0.054)
Opp. can finish with 2 darts			-0.014			-0.011
			(0.053)			(0.054)
Opp. can finish with 3 darts			0.045			0.062
			(0.058)			(0.059)
Decisive leg x Opp. can finish					-0.190	
					(0.194)	
Decisive leg x Opp. can finish with 1 dart						-0.254
						(0.207)
Decisive leg x Opp. can finish with 2 darts						-0.064
						(0.212)
Decisive leg x Opp. can finish with 3 darts						-0.272
						(0.239)
Panel D: Youth						
Decisive leg	-0.278^{***}			-0.149	0.034	0.038
	(0.064)			(0.099)	(0.251)	(0.251)
Opp. can finish		-0.169^{**}			-0.148^{*}	
		(0.079)			(0.084)	
Opp. can finish with 1 dart			-0.238^{***}			-0.205^{**}
			(0.091)			(0.096)
Opp. can finish with 2 darts			-0.159^{*}			-0.159^{*}
			(0.089)			(0.094)
Opp. can finish with 3 darts			-0.106			-0.104
			(0.095)			(0.100)
Decisive leg x Opp. can finish					-0.209	
					(0.263)	
Decisive leg x Opp. can finish with 1 dart						-0.352
						(0.281)
Decisive leg x Opp. can finish with 2 darts						-0.085
						(0.290)
Decisive leg x Opp. can finish with 3 darts						-0.055
						(0.327)

Table 5.6: Coefficient estimates underlying the average marginal effects in Table 5.3

	Coefficient	z-value	p-value
Panel A: International			
1/2/3 darts	-0.034	-0.413	0.680
1 dart	-0.129	-1.136	0.256
2 darts	0.022	0.175	0.861
3 darts	0.094	0.544	0.587
Panel B: BICC			
1/2/3 darts	-0.287	-8.659	0.000
1 dart	-0.396	-9.745	0.000
2 darts	-0.215	-4.988	0.000
3 darts	-0.158	-2.761	0.006
Panel C: Super League			
1/2/3 darts	-0.170	-1.964	0.050
1 dart	-0.227	-2.000	0.046
2 darts	-0.083	-0.693	0.488
3 darts	-0.217	-1.355	0.175
Panel D: Youth			
1/2/3 darts	-0.323	-2.529	0.011
1 dart	-0.519	-3.237	0.001
2 darts	-0.206	-1.204	0.228
3 darts	-0.121	-0.541	0.588

.

Table 5.7: Regression results for the likelihood of finishing, controlling for points-left fixed effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel A: International						
Decisive leg	-0.021^{**}			-0.008	0.022	0.022
	(0.008)			(0.008)	(0.026)	(0.026)
Opp. can finish		0.002			0.003	
		(0.005)			(0.005)	
Opp. can finish with 1 dart			-0.003			-0.001
			(0.005)			(0.006)
Opp. can finish with 2 darts			0.010^{*}			0.012^{**}
			(0.005)			(0.006)
Opp. can finish with 3 darts			-0.002			-0.002
			(0.006)			(0.006)
Decisive leg x Opp. can finish					-0.033	
					(0.028)	
Decisive leg x Opp. can finish with 1 dart						-0.044
						(0.029)
Decisive leg x Opp. can finish with 2 darts						-0.033
						(0.030)
Decisive leg x Opp. can finish with 3 darts						-0.005
						(0.033)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Points-left fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	3	3	3	3	3	3
McFadden pseudo R-squared	0.151	0.094	0.094	0.094	0.094	0.094
Observations	111,241	68,547	68,547	68,547	68,547	68,547
Effective observations	84,994	37,837	37,837	37,837	37,837	37,837
Panel B: BICC						
Decisive leg	-0.072^{***}			-0.057^{***}	-0.007	-0.008
	(0.003)			(0.003)	(0.008)	(0.008)
Opp. can finish		0.005^{**}			0.010^{***}	
		(0.002)			(0.003)	
Opp. can finish with 1 dart			0.005^{**}			0.011^{***}
			(0.003)			(0.003)
Opp. can finish with 2 darts			0.005^{**}			0.008^{***}
			(0.003)			(0.003)
Opp. can finish with 3 darts			0.006^{**}			0.008^{***}
			(0.003)			(0.003)
Decisive leg x Opp. can finish					-0.056^{***}	
					(0.008)	
Decisive leg x Opp. can finish with 1 dart						-0.076^{***}
						(0.009)
Decisive leg x Opp. can finish with 2 darts						-0.041^{***}
						(0.009)
Decisive leg x Opp. can finish with 3 darts						-0.030^{***}
						(0.010)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Points-left fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	4	3	3	3	3	3
McFadden pseudo R-squared	0.135	0.079	0.079	0.080	0.080	0.080
Observations	587,876	345,416	345,416	345,416	345,416	345,416
Effective observations	463,690	193,135	193,135	193,135	193,135	193,135

Notes: The table reports the average marginal effects resulting from logit regression analyses of finishing performance across throws where the player can finish the leg with one dart. Model 1 uses all one-dart finish opportunities, whereas Models 2-6 use players' first one-dart finish opportunity in a leg only. Definitions are as in Table 5.2. The regression specifications now in addition include points-left fixed effects.

Table 5.7: Continued

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel C: Super League						
Decisive leg	-0.032^{***}			-0.036^{***}	-0.002	-0.003
	(0.009)			(0.009)	(0.023)	(0.023)
Opp. can finish		0.003			0.006	
		(0.006)			(0.006)	
Opp. can finish with 1 dart			0.006			0.009
			(0.007)			(0.007)
Opp. can finish with 2 darts			-0.003			-0.003
			(0.007)			(0.007)
Opp. can finish with 3 darts			0.009			0.012
			(0.007)			(0.008)
Decisive leg x Opp. can finish					-0.038	
					(0.024)	
Decisive leg x Opp. can finish with 1 dart						-0.053^{**}
						(0.026)
Decisive leg x Opp. can finish with 2 darts						-0.014
						(0.026)
Decisive leg x Opp. can finish with 3 darts						-0.048
						(0.030)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Points-left fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	2	3	3	3	3	3
McFadden pseudo R-squared	0.139	0.085	0.085	0.085	0.085	0.085
Observations	81,385	46,518	46,518	46,518	46,518	46,518
Effective observations	65,063	26,536	26,536	26,536	26,536	26,536
Panel D: Youth						
Decisive leg	-0.040^{***}			-0.028^{***}	0.007	0.006
	(0.010)			(0.010)	(0.026)	(0.025)
Opp. can finish		-0.031^{***}			-0.027^{***}	
		(0.009)			(0.009)	
Opp. can finish with 1 dart			-0.042^{***}			-0.034^{***}
			(0.010)			(0.010)
Opp. can finish with 2 darts			-0.029^{***}			-0.028^{***}
			(0.009)			(0.010)
Opp. can finish with 3 darts			-0.021^{**}			-0.020^{**}
			(0.010)			(0.010)
Decisive leg x Opp. can finish					-0.040	
					(0.026)	
Decisive leg x Opp. can finish with 1 dart						-0.068^{**}
						(0.028)
Decisive leg x Opp. can finish with 2 darts						-0.016
						(0.029)
Decisive leg x Opp. can finish with 3 darts						-0.008
						(0.033)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Points-left fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	3	3	3	3	3	3
McFadden pseudo R-squared	0.151	0.074	0.074	0.073	0.074	0.075
Observations	45,893	20,563	20,563	20,563	20,563	20,563
Effective observations	36,385	9,322	9,322	9,322	9,322	9,322

Table 5.8: Regression results for the likelihood of finishing, using leg-format International matches only

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Decisive leg	-0.021^{**}			0.004	0.060^{**}	0.060^{*}
	(0.010)			(0.010)	(0.030)	(0.030)
Opp. can finish		0.002			0.004	
		(0.005)			(0.005)	
Opp. can finish with 1 dart			-0.005			-0.002
			(0.006)			(0.006)
Opp. can finish with 2 darts			0.013^{**}			0.014^{**}
			(0.006)			(0.006)
Opp. can finish with 3 darts			-0.002			-0.001
			(0.007)			(0.007)
Decisive leg x Opp. can finish					-0.063^{*}	
					(0.033)	
Decisive leg x Opp. can finish with 1 dart						-0.077^{**}
						(0.034)
Decisive leg x Opp. can finish with 2 darts						-0.058
						(0.036)
Decisive leg x Opp. can finish with 3 darts						-0.031
						(0.040)
Data	All	First only	First only	First only	First only	First only
Player-match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial throw number (order)	3	1	1	1	1	1
McFadden pseudo R-squared	0.142	0.088	0.088	0.088	0.088	0.089
Observations	100,292	59,466	59,466	59,466	59,466	59,466
Effective observations	79,565	34,688	34,688	34,688	34,688	34,688

Notes: The table reports the average marginal effects resulting from logit regression analyses of finishing performance across throws where the player can finish the leg with one dart, for International matches with a leg format only. Model 1 uses all one-dart finish opportunities, whereas Models 2-6 use players' first one-dart finish opportunity in a leg only. Definitions are as in Table 5.2.

Table 5.9: Regression results for the number of points thrown, using leg-format International matches only

	Model 1: Throw 1	Model 2: Throw 2	Model 3: Throw 3	Model 4: Throws 1-3
Decisive leg	-0.277	-0.803	-1.077^{*}	-0.711^{**}
	(0.611)	(0.612)	(0.609)	(0.350)
Player-match fixed effects	Yes	Yes	Yes	Yes
Throw fixed effects	-	-	-	Yes
Polynomial throw number (order)	7	4	3	7
Observations	142,838	142,838	142,831	428,507
Adjusted R-squared	0.240	0.242	0.236	0.250

Notes: The table reports the coefficients resulting from OLS regression analyses of the number of points thrown in the first, second, and/or third throw of a leg, for International matches with a leg format only. Definitions are as in Table 5.4.

Summary

This dissertation uses four naturally occurring data sets to study psychological factors that influence various real-life decisions.

Chapter 2 examines the optimality of strategic decision making in the *Showcase Showdown*, a simple sequential game of perfect information in the American TV show *The Price is Right*. We show that contestants often deviate from the game-theoretic optimum. These deviations can neither be explained by random decision errors nor by a preference for harm caused by inaction over the equivalent harm caused by an explicit action. Instead, the observed behavior can be explained by limited foresight, where a contestant only thinks ahead to the next stage of the game. This finding illustrates that even a relatively simple environment with high stakes and a publicly available optimal strategy may not be sufficient to ensure the descriptive validity of game-theoretic predictions.

Chapter 3 studies the effect of marginally trailing on performance in professional sports matches. We extend Berger and Pope's (2011) analysis of whether marginally trailing improves the odds of winning in basketball to Australian football, American football and rugby, and find little to no evidence of such an effect for these three sports. When we revisit the phenomenon for basketball, we do find supportive evidence for National Basketball Association (NBA) matches from the period analyzed in Berger and Pope. However, we find no significant effect for NBA matches from outside this sample period, for collegiate matches, and for matches from the Women's NBA. Moreover, our high-powered meta-analyses across the different sports and competitions cannot reject the hypothesis of no effect of marginally trailing on winning, and the confidence intervals suggest that the true effect, if existent at all, is likely relatively small. Our results suggest that the performance-enhancing effect of trailing in a competition may disappear when agents are experienced and when they receive continuous feedback on the score difference.

Chapter 4 explores the role of transaction utility in prescription medicine purchases. We exploit a unique data set containing medicine purchases at 85 percent of US pharmacies to estimate the causal effect of discounts on the demand for life-saving drugs. We demonstrate that discounts, in the form of copay assistance, have a significant and positive effect on patients' propensity to purchase their drugs at a given price. Since we identify the effect for a fixed price, the increased demand indicates a shift of, rather than a movement along, the demand curve. Our results can be explained by the transaction utility introduced by discounts.

Chapter 5 investigates how within-match variation in incentives affects the performance of darts players. We analyze four data sets covering a total of 29,381 darts matches of professional, amateur, and youth players, and find that amateur and youth players display a sizable performance decrease at decisive moments. Professional players appear less susceptible of such choking under pressure.

Summary in Dutch

Dit proefschrift maakt gebruik van vier natuurlijke datasets om te onderzoeken welke gedragseconomische factoren van invloed zijn op diverse beslissingen in het echte leven.

Hoofdstuk 2 onderzoekt de optimaliteit van strategische beslissingen van deelnemers aan de *Showcase Showdown*, een simpel, sequentieel spel met complete informatie in de Amerikaanse TV show *The Price is Right*. Onze resultaten laten zien dat deelnemers regelmatig afwijken van de optimale strategie. We onderzoeken of deze afwijkingen kunnen worden verklaard door willekeurige misrekeningen of door een voorkeur voor negatieve consequenties die voortkomen uit inactie ten opzichte van dezelfde consequenties voortkomend uit expliciete actie, maar wijzen uiteindelijk beide verklaringen af. Het geobserveerde gedrag kan wél worden verklaard door de de aanwezigheid van een groot aantal spelers dat het spel simplificeert door slechts één stap vooruit te denken. Onze bevindingen laten zien dat speltheoretische voorspellingen menselijk gedrag zelfs niet goed beschrijven in een relatief simpele setting waar veel geld op het spel staat en waar de optimale strategie online beschikbaar is.

Hoofdstuk 3 onderzoekt het effect van een kleine achterstand op het prestatieniveau van teams in sportwedstrijden. Berger and Pope (2011) hebben laten zien dat een lichte achterstand de winkans vergroot in basketbal, zowel op professioneel als op universiteitsniveau. We breiden hun analyse uit naar Australian football, American football en rugby. We vinden echter weinig tot geen bewijs van een dergelijk effect voor deze drie sporten. Wanneer we het fenomeen opnieuw onderzoeken in basketbal vinden we ondersteunend bewijs voor wedstrijden uit de National Basketball Association (NBA) die plaats hebben gevonden binnen de onderzoeksperiode van Berger en Pope. We vinden echter geen significant effect voor NBA wedstrijden die buiten deze periode plaats hebben gevonden, noch voor universiteitswedstrijden en wedstrijden uit de WNBA. Onze meta-analyses die de verschillende sporten en competities combineren kunnen de nulhypothese niet verwerpen dat er geen effect is van een lichte achterstand op de winstkans.

Hoofdstuk 4 onderzoekt de rol van transactienut bij het kopen van geneesmiddelen. We gebruiken een unieke dataset waarin de medicijnaankopen bij 85 procent van alle Amerikaanse apotheken staan geregistreerd om het causale verband van kortingen op de kans dat patiënten hun medicijnen kopen te schatten. In overeenstemming met het idee van transactienut zien we dat kortingen de vraagcurve verschuiven, wat ertoe leidt dat patiënten vaker hun voorgeschreven medicijnen kopen, zelfs als de uiteindelijke prijs niet verandert.

Hoofdstuk 5 onderzoekt hoe variatie in prestatieprikkels binnen dartwedstrijden de prestaties van darters beïnvloedt. We analyseren vier datasets met in totaal 29.381 dartwedstrijden van professionele, amateur- en jeugdspelers. We zien dat amateur- en jeugdspelers op beslissende momenten een forse prestatieafname laten zien. Professionele spelers blijken onder druk minder vatbaar voor ‘choking under pressure’.

Bibliography

Abadie, A. (2020), 'Statistical non-significance in empirical economics', *American Economic Review: Insights* **2**(2), 193–208.

Abadie, A. and Imbens, G. W. (2011), 'Bias-corrected matching estimators for average treatment effects', *Journal of Business and Economic Statistics* **29**(1), 1–11.

Allen, E. J., Dechow, P. M., Pope, D. G. and Wu, G. (2016), 'Reference-dependent preferences: Evidence from marathon runners', *Management Science* **63**(6), 1657–1672.

Anderson, S. P., Goeree, J. K. and Holt, C. A. (2001), 'Minimum-effort coordination games: Stochastic potential and logit equilibrium', *Games and Economic Behavior* **34**(2), 177–199.

Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009), 'Large stakes and big mistakes', *Review of Economic Studies* **76**(2), 451–469.

Baghestanian, S. and Frey, S. (2016), 'GO figure: Analytic and strategic skills are separable', *Journal of Behavioral and Experimental Economics* **64**, 71–80.

Baltussen, G., van den Assem, M. J. and van Dolder, D. (2016), 'Risky choice in the limelight', *Review of Economics and Statistics* **98**(2), 318–332.

Baumeister, R. F. (1984), 'Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance', *Journal of Personality and Social Psychology* **46**(3), 610–620.

Baumeister, R. F. and Showers, C. J. (1986), 'A review of paradoxical performance effects: Choking under pressure in sports and mental tests', *European Journal of Social Psychology* **16**(4), 361–383.

Becker, B. E. and Huselid, M. A. (1992), 'The incentive effects of tournament compensation systems', *Administrative Science Quarterly* **37**(2), 336–350.

Beilock, S. L. (2010), *Choke: What the Secrets of the Brain Reveal About Getting It Right When You Have To*, New York: Free Press.

Beilock, S. L. and Carr, T. H. (2001), 'On the fragility of skilled performance: What governs choking under pressure?', *Journal of Experimental Psychology: General* **130**(4), 701–725.

Belot, M., Bhaskar, V. and van de Ven, J. (2010), 'Promises and cooperation: Evidence from a TV game show', *Journal of Economic Behavior and Organization* **73**(3), 396–405.

Bennett, R. W. and Hickman, K. A. (1993), 'Rationality and the 'Price is Right'', *Journal of Economic Behavior and Organization* **21**(1), 99–105.

Berchick, E. R., Hood, E. and Barnett, J. C. (2018), 'Health insurance coverage in the United States: 2017', *Current Population Reports. Washington DC: US Government Printing Office.* .

Berger, J. and Pope, D. G. (2011), 'Can losing lead to winning?', *Management Science* **57**(5), 817–827.

Berk, J. B., Hughson, E. and Vandezande., K. (1996), 'The Price is Right, but are the bids? An investigation of rational decision theory', *American Economic Review* **86**(4), 954–970.

Berkowitz, E. N. and Walton, J. R. (1980), 'Contextual influences on consumer price responses: An experimental analysis', *Journal of Marketing* **3**(3), 349–358.

Binmore, K. (1999), 'Why experiment in economics?', *Economic Journal* **109**(453), F16–F24.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L. and Shaked, A. (2002), 'A backward induction experiment', *Journal of Economic Theory* **104**(1), 48–88.

Böheim, R., Grübl, D. and Lackner, M. (2019), 'Choking under pressure – Evidence of the causal effect of audience size on performance', *Journal of Economic Behavior and Organization* **168**, 76–93.

Cai, H. and Wang, J. T.-Y. (2006), 'Overcommunication in strategic information transmission games', *Games and Economic Behavior* **56**(1), 7–36.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014), 'Robust nonparametric confidence intervals for regression discontinuity designs', *Econometrica* **82**(6), 2295–2326.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2015), 'Optimal data-driven regression discontinuity plots', *Journal of the American Statistical Association* **110**(512), 1753–1769.

Camerer, C. F. (2015), 'The promise and success of lab–field generalizability in experimental economics: A critical reply to levitt and list'. *In: Guillaume R. Fréchette, Andrew Schotter (Eds.), Handbook of Experimental Economic Methodology.*

Camerer, C. F. and Hogarth, R. M. (1999), 'The effects of financial incentives in experiments: A review and capital-labor-production framework', *Journal of Risk and Uncertainty* **19**, 7–42.

Campbell, M. C. (1999), 'Perceptions of price unfairness: Antecedents and consequences', *Journal of Marketing Research* **36**(2), 187–199.

Cao, Z., Price, J. and Stone, D. F. (2011), 'Performance under pressure in the NBA', *Journal of Sports Economics* **12**(3), 231–252.

Capra, C. M., Goeree, J. K., Gomez, R. and Holt, C. A. (1999), 'Anomalous behavior in a traveler's dilemma?', *American Economic Review* **89**(3), 678–690.

Carlin, B. I. and Robinson, D. T. (2009), 'Fear and loathing in Las Vegas: Evidence from blackjack tables', *Judgment and Decision Making* **4**(5), 385–396.

Cattaneo, M. D., Titiunik, R. and Vazquez-Bare, G. (2019), 'Power calculations for regression discontinuity designs', *Stata Journal* **19**(1), 210–245.

Charness, G. and Rabin, M. (2002), 'Understanding social preferences with simple tests', *Quarterly Journal of Economics* **117**(3), 817–869.

Chiappori, P. A., Levitt, S. D. and Groseclose, T. (2002), 'Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer', *American Economic Review* **92**(4), 1138–1151.

Coe, P. R. and Butterworth, W. (1995), 'Optimal stopping in "the Showcase Showdown"', *The American Statistician* **49**(3), 271–275.

Cohen-Zada, D., Krumer, A. and Shtudiner, Z. (2017), 'Psychological momentum and gender', *Journal of Economic Behavior and Organization* **135**, 66–81.

Corgnet, B., Gómez-Miñambres, J. and Hernán-González, R. (2015), 'Goal setting and monetary incentives: When large stakes are not enough', *Management Science* **61**(12), 2926–2944.

Dafny, L., Ody, C. and Schmitt, M. (2017), 'When discounts raise costs: The effect of copay coupons on generic utilization', *American Economic Journal: Economic Policy* **9**(2), 91–123.

Daniel, H. and Bornstein, S. S. (2019), 'Policy recommendations for pharmacy benefit managers to stem the escalating costs of prescription drugs: A position paper from the American college of physicians.', *Annals of Internal Medicine* **171**(11), 823–824.

Dechenaux, E., Kovenock, D. and Sheremeta, R. (2015), 'A survey of experimental research on contests, all-pay auctions and tournaments', *Experimental Economics* **18**(14), 609–669.

Dixit, A. (1982), 'Recent developments in oligopoly theory', *American Economic Review* **72**(2), 12–17.

Dohmen, T. J. (2008), 'Do professionals choke under pressure?', *Journal of Economic Behavior and Organization* **65**(3-4), 636–653.

Dufwenberg, M., Sundaram, R. and Butler, D. J. (2010), 'Epiphany in the game of 21', *Journal of Economic Behavior and Organization* **75**(2), 132–143.

Dufwenberg, M. and Van Essen, M. (2018), 'King of the hill: Giving backward induction its best shot', *Games and Economic Behavior* **112**, 125–138.

Ehrenberg, R. G. and Bognanno, M. L. (1990*a*), 'Do tournaments have incentive effects?', *Journal of Political Economy* **98**(6), 1307–1324.

Ehrenberg, R. G. and Bognanno, M. L. (1990*b*), 'The incentive effects of tournaments revisited: Evidence from the European PGA tour', *Industrial and Labor Relations Review* **43**(3), 74–88.

Falk, A. and Heckman, J. J. (2009), 'Lab experiments are a major source of knowledge in the social sciences', *Science* **326**(5952), 535–538.

Fehr, E. and Gachter, S. (2000), 'Cooperation and punishment in public goods experiments', *American Economic Review* **90**(4), 980–994.

Fehr, E. and Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *Quarterly Journal of Economics* **114**(3), 817–868.

Fernández-Val, I. (2009), 'Fixed effects estimation of structural parameters and marginal effects in panel probit models', *Journal of Econometrics* **150**(1), 71–85.

Fershtman, C. and Gneezy, U. (2011), 'The tradeoff between performance and quitting in high power tournaments', *Journal of the European Economic Association* **9**(2), 318–336.

Fey, M., McKelvey, R. D. and Palfrey, T. P. (1996), 'An experimental study of constant-sum centipede games', *International Journal of Game Theory* **25**(3), 269–287.

Frederick, S., Loewenstein, G. and O'Donoghue, T. (2002), 'Time discounting and time preference: A critical review', *Journal of Economic Literature* **40**(2), 351–401.

Gabaix, X. and Laibson, D. (2005), Bounded rationality and directed cognition. Working Paper.

Gabaix, X., Laibson, D., Moloche, G. and Weinberg, S. (2006), 'Costly information acquisition: Experimental analysis of a boundedly rational model', *American Economic Review* **96**(4), 1043–1068.

Gauriot, R. and Page, L. (2015), 'I take care of my own: A field study on how leadership handles conflict between individual and collective incentives', *American Economic Review* **105**(5), 414–419.

Gauriot, R. and Page, L. (2019), 'Does success breed success? A quasi-experiment on strategic momentum in dynamic contests', *Economic Journal* **129**(624), 3107–3136.

Gertner, R. (1993), 'Game shows and economic behavior: Risk-taking on card sharks', *Quarterly Journal of Economics* **108**(2), 507–521.

Gill, D. and Prowse, V. (2012), 'A structural analysis of disappointment aversion in a real effort competition', *American Economic Review* **102**(1), 469–503.

Gilsdorf, K. F. and Sukhatme, V. A. (2008*a*), 'Testing Rosen's sequential elimination tournament model: Incentives and player performance in professional tennis', *Journal of Sports Economics* **9**(3), 287–303.

Gilsdorf, K. F. and Sukhatme, V. A. (2008*b*), 'Tournament incentives and match outcomes in women's professional tennis', *Applied Economics* **40**(18), 2405–2412.

Gneezy, A., Gneezy, U. and Lauga, D. O. (2014), 'A reference-dependent model of the price–quality heuristic', *Journal of Marketing Research* **51**(2), 153–164.

Gneezy, U. and Rustichini, A. (2000), 'Pay enough or don't pay at all', *Quarterly Journal of Economics* **115**(3), 791–810.

Gneezy, U., Rustichini, A. and Vostroknutov, A. (2010), 'Experience and insight in the race game', *Journal of Economic Behavior and Organization* **75**(2), 144–155.

Goeree, J. K., Holt, C. A. and Palfrey, T. R. (2002), 'Quantal response equilibrium and overbidding in private-value auctions', *Journal of Economic Theory* **104**(1), 247–272.

Goeree, J. K., Holt, C. A. and Palfrey, T. R. (2003), 'Risk averse behavior in generalized matching pennies games', *Games and Economic Behavior* **45**(1), 97–113.

Goldman, M. and Rao, J. M. (2012), 'Effort vs. concentration: The asymmetric impact of pressure on NBA performance', *Proceedings of the MIT Sloan Sports Analytics Conference 2012*.

Grewal, D., Krishnan, R., Baker, J. and Borin, N. A. (1998), 'The effect of store name, brand name and price discounts on consumers' evaluations and purchase intentions', *Journal of Retailing* **74**(3), 331–352.

Grosjean, J. H. (1998), 'Beating the Showcase Showdown', *Chance* **11**(1), 14–19.

Hahn, J., Todd, P. and Van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression discontinuity design', *Econometrica* **69**(1), 201–209.

Harb-Wu, K. and Krumer, A. (2019), 'Choking under pressure in front of a supportive audience: Evidence from professional biathlon', *Journal of Economic Behavior and Organization* **166**, 246–262.

Harris, C. and Vickers, J. (1987), 'Racing with uncertainty', *Review of Economic Studies* **54**(1), 1–21.

Heath, C., Larrick, R. P. and Wu, G. (1999), 'Goals as reference points', *Cognitive Psychology* **38**(1), 79–109.

Hedges, L. V. and Vevea, J. L. (1998), 'Fixed- and random-effects models in meta-analysis', *Psychological Methods* **3**(4), 486–504.

Hickman, D. C. and Metz, N. E. (2015), 'The impact of pressure on performance: Evidence from the PGA tour', *Journal of Economic Behavior and Organization* **116**, 319–330.

Ho, T. H. and Su, X. (2013), 'A dynamic level-k model in sequential games', *Management Science* **59**(2), 452–469.

Huang, J. (2018), The thrill of the deal: Quantifying the price of perceived discounts and mark-ups. Working paper.

Isoni, A. (2011), 'The willingness-to-accept/willingness-to-pay disparity in repeated markets: Loss aversion or 'bad-deal'aversion?', *Theory and Decision* **71**(3), 409–430.

Jackson, D. and Turner, R. (2017), 'Power analysis for random-effects meta-analysis', *Research Synthesis Methods* **8**(3), 290–302.

Jehiel, P. (1995), ‘Limited horizon forecast in repeated alternate games’, *Journal of Economic Theory* **67**(2), 497–519.

Jehiel, P. (1998), ‘Learning to play limited forecast equilibria’, *Games and Economic Behavior* **22**(2), 274–298.

Johnson, E. J., Camerer, C. F., Sen, S. and Rymon, T. (2002), ‘Detecting failures of backward induction: Monitoring information search in sequential bargaining’, *Journal of Economic Theory* **104**(1), 16–47.

Kahneman, D. and Tverky, A. (1979), ‘Prospect theory: An analysis of decision under risk’, *Econometrica* **47**(2), 263–291.

Kamenica, E. (2012), ‘Behavioral economics and psychology of incentives’, *Annual Review of Economics* **4**(1), 427–452.

Ke, S. (2019), ‘Boundedly rational backward induction’, *Theoretical Economics* **14**(1), 103–134.

Keren, G. B. and Wagenaar, W. A. (1985), ‘On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory’, *Journal of Experimental Psychology: General* **114**(2), 133–158.

King, G. J., Chao, X. and Duenyas, I. (2019), ‘Who benefits when prescription drug manufacturers offer copay coupons?’, *Management Science* **8**(65), 3758–3775.

Klumpp, T. and Polborn, M. K. (2006), ‘Primaries and the New Hampshire effect’, *Journal of Public Economics* **90**(6-7), 1073–1114.

Konrad, K. A. (2009), *Strategy and Dynamics in Contests*, Oxford University Press.

Lazear, E. P. (2000), ‘Performance pay and productivity’, *American Economic Review* **90**(5), 1346–1361.

Lee, C.-Y. (2020), ‘Pricing strategy and moral hazard: Copay coupons in pharmaceuticals’, *International Journal of Industrial Organization* **70**, 102611.

Lee, D. S. (2008), 'Randomized experiments from non-random selection in US house elections', *Journal of Econometrics* **142**(2), 675–697.

Levitt, S. D. (2004), 'Testing theories of discrimination: Evidence from Weakest Link', *Journal of Law and Economics* **47**(2), 431–453.

Levitt, S. D. and List, J. A. (2007*a*), 'Viewpoint: On the generalizability of lab behaviour to the field', *Canadian Journal of Economics* **40**(2), 347–370.

Levitt, S. D. and List, J. A. (2007*b*), 'What do laboratory experiments measuring social preferences reveal about the real world?', *Journal of Economic Perspectives* **21**(2), 153–174.

Levitt, S. D. and List, J. A. (2008), 'Homo economicus evolves', *Science* **319**(5865), 909–910.

Levitt, S. D., List, J. A., Neckermann, S. and Sadoff, S. (2016), 'The behavioralist goes to school: Leveraging behavioral economics to improve educational performance', *American Economic Journal: Economic Policy* **8**(4), 183–219.

Levitt, S. D., List, J. A. and Sadoff, S. E. (2011), 'Checkmate: Exploring backward induction among chess players', *American Economic Review* **101**(2), 975–990.

Lindner, F. (2017), Choking under pressure of top performers: Evidence from biathlon competitions. Working paper.

List, J. A. (2006), 'Friend or Foe? A natural experiment of the prisoner's dilemma', *Review of Economics and Statistics* **88**(3), 463–471.

Loewenstein, G. and Prelec, D. (1992), 'Anomalies in intertemporal choice: Evidence and an interpretation', *Quarterly Journal of Economics* **107**(2), 573–597.

Lynch, J. G. (2005), 'The effort effects of prizes in the second half of tournaments', *Journal of Economic Behavior and Organization* **57**(1), 115–129.

Malueg, D. A. and Yates, A. J. (2010), 'Testing contest theory: Evidence from best-of-three tennis matches', *Review of Economics and Statistics* **92**(3), 689–692.

Mantovani, M. (2015), Limited backward induction: Foresight and behavior in sequential games. Working Paper, University of Milano-Bicocca, Department of Economics.

Masters, R. and Maxwell, J. (2008), 'The theory of reinvestment', *International Review of Sport and Exercise Psychology* **1**(2), 160–183.

McKelvey, R. D. and Palfrey, T. R. (1992), 'An experimental study of the centipede game', *Econometrica* **60**(4), 803–836.

McKelvey, R. D. and Palfrey, T. R. (1998), 'Quantal response equilibria for extensive form games', *Experimental Economics* **1**(1), 9–41.

Metrick, A. (1995), 'A natural experiment in 'Jeopardy'', *American Economic Review* **85**(1), 240–253.

Moore, D. J. and Olshavsky, R. W. (1989), 'Brand choice and deep price discounts', *Psychology and Marketing* **6**(3), 181–196.

Moskowitz, T. J. and Wertheim, L. J. (2011), *Scorecasting: The hidden influences behind how sports are played and games are won*, New York: Crown Pub, chapter 2, pp. 7–30.

Muehlbacher, S., Kirchler, E. and Kunz, A. (2015), 'The impact of transaction utility on consumer decisions', *Zeitschrift Für Psychologie* **219**, 217–223.

Oberholzer-Gee, F., Waldfogel, J. and White, M. W. (2010), 'Friend or Foe? Cooperationand learning in high-stakes games', *Review of Economics and Statistics* **92**(1), 179–187.

OECD (2020), 'Main economic indicators - complete database', Accessed on 4 March 2019. Main Economic Indicators (database).

Ötting, M., Deutscher, C., Schneemann, S., Langrock, R., Gehrmann, S. and Scholten, H. (2020), 'Performance under pressure in skill tasks: An analysis of professional darts', *PLoS ONE* **15**(2), e0228870.

Paarsch, H. and Shearer, B. S. (1999), 'The response of worker effort to piece rates: Evidence from the British Columbia tree-planting industry', *Journal of Human Resources* **34**(4), 643–667.

Page, L. and Coates, J. (2017), 'Winner and loser effects in human competitions. Evidence from equally matched tennis players', *Evolution and Human Behavior* **38**, 530–535.

Palacios-Huerta, I. and Volij, O. (2009), 'Field centipedes', *American Economic Review* **99**(4), 1619–1635.

Panityakul, T., Bumrungsup, C. and Knapp, G. (2013), 'On estimating residual heterogeneity in random-effects meta-regression: A comparative study', *Journal of Statistical Theory and Applications* **12**(3), 253–265.

Paserman, M. D. (2010), Gender differences in performance in competitive environments? Evidence from professional tennis players. Working paper.

Paule, R. C. and Mandel, J. (1989), 'Consensus values, regressions, and weighting factors', *Journal of Research of the National Institute of Standards and Technology* **94**(3), 197–203.

Pope, D. G. and Simonsohn, U. (2011), 'Round numbers as goals: Evidence from baseball, SAT takers and the lab', *Psychological Science* **22**(1), 71–79.

Post, T., van den Assem, M. J., Baltussen, G. and Thaler, R. H. (2008), 'Deal or no deal? Decision making under risk in a large-payoff game show', *American Economic Review* **98**(1), 38–71.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37**(1), 7–63.

Price, J. and Wolfers, J. (2010), 'Racial discrimination among NBA referees', *Quarterly Journal of Economics* **125**(4), 1859–1887.

Rampal, J. (2018), Limited foresight equilibrium. Working paper, Indian Institute of Management Ahmedabad.

Rao, A. R. and Monroe, K. B. (1989), 'The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review', *Journal of Marketing Research* **26**(3), 351–357.

Rapoport, A., Stein, W. E., Parco, J. E. and Nicholas, T. E. (2003), 'Equilibrium play and adaptive learning in a three-person centipede game', *Games and Economic Behavior* **43**(2), 239–265.

Ritov, I. and Baron, J. (1990), 'Reluctance to vaccinate: Omission bias and ambiguity', *Journal of Behavioral Decision Making* **3**(4), 263–277.

Rosenthal, R. W. (1981), 'Games of perfect information, predatory pricing and the chain-store paradox', *Journal of Economic Theory* **25**(1), 92–100.

Royce, T. J., Kircher, S. and Conti, R. M. (2019), 'Pharmacy benefit manager reform: Lessons from Ohio', *Jama* **322**(4), 299–300.

Rubinstein, A. (1982), 'Perfect equilibrium in a bargaining model', *Econometrica* **50**(1), 97–109.

Selten, R. (1978), 'The chain store paradox', *Theory and Decision* **9**, 127–159.

Shearer, B. (2004), 'Piece rates, fixed wages and incentives: Evidence from a field experiment', *Review of Economic Studies* **71**(2), 513–534.

Shiv, B., Carmon, Z. and Ariely, D. (2005), 'Placebo effects of marketing actions: Consumers may get what they pay for', *Journal of Marketing Research* **42**(4), 383–393.

Sinha, I. and Smith, M. F. (2000), 'Consumers' perceptions of promotional framing of price', *Psychology and Marketing* **17**(3), 257–275.

Smith, V. L. and Walker, J. M. (1993), 'Monetary rewards and decision cost in experimental economics', *Economic Inquiry* **31**(2), 245–261.

Spenkuch, J. L., Montagnes, B. P. and Magleby, D. B. (2018), 'Backward induction in the wild? Evidence from sequential voting in the US senate', *American Economic Review* **108**(7), 1971–2013.

Spranca, M., Minsk, E. and Baron, J. (1991), 'Omission and commission in judgment and choice', *Journal of Experimental Social Psychology* **27**(1), 76–105.

Stigler, G. J. (1961), 'The economics of information', *Journal of Political Economy* **69**(3), 213–225.

Tenorio, R. and Cason, T. N. (2002), 'To spin or not to spin? Natural and laboratory experiments from The Price is Right', *Economic Journal* **112**, 170–195.

Thaler, R. H. (1981), 'Some empirical evidence on dynamic inconsistency', *Economics Letters* **8**(3), 201–207.

Thaler, R. H. (1985), 'Mental accounting and consumer choice', *Marketing Science* **4**(3), 199–214.

Thaler, R. H. (1999), 'Mental accounting matters', *Journal of Behavioral Decision Making* **12**(3), 183–206.

Thistlethwaite, D. L. and Campbell, D. T. (1960), 'Regression discontinuity analysis: An alternative to the ex post facto experiment', *Journal of Educational Psychology* **51**(6), 309–317.

Tibshirani, R. J., Price, A. and Taylor, J. (2011), 'A statistician plays darts', *Journal of the Royal Statistical Society: Series A* **174**(1), 213–226.

Toma, M. (2017), 'Missed shots at the free-throw line', *Journal of Sports Economics* **18**(6), 185–207.

Turmunkh, U., van den Assem, M. J. and van Dolder, D. (2019), 'Malleable lies: Communication and cooperation in a high stakes TV game show', *Management Science* **65**(10), 4795–4812.

Tversky, A. and Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**(4), 297–323.

van den Assem, M. J., van Dolder, D. and Thaler, R. H. (2012), 'Split or Steal? Cooperative behavior when the stakes are large', *Management Science* **58**(1), 2–20.

van Dolder, D., van den Assem, M. J., Camerer, C. F. and Thaler, R. H. (2015), 'Standing united or falling divided? High stakes bargaining in a TV game show', *American Economic Review* **105**(5), 402–407.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D. and Salanti, G. (2015), 'Methods to estimate the between-study variance and its uncertainty in meta-analysis', *Research Synthesis Methods* **7**(1), 55–79.

von Stackelberg, H. (1934), *Marktform und Gleichgewicht*, Vienna: Springer.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**(2), 307–333.

Walker, J., Risen, J. L., Gilovich, T. and Thaler, R. (2018), 'Sudden death aversion: Avoiding superior options because they feel riskier', *Journal of Personality and Social Psychology* **115**(3), 363–378.

Weaver, R. and Frederick, S. (2012), 'A reference price theory of the endowment effect', *Journal of Marketing Research* **49**(5), 696–707.

Yerkes, R. M. and Dodson, J. D. (1908), 'The relationship of strength of stimulus to rapidity of habit formation', *Journal of Comparative Neurology and Psychology* **18**, 459–482.

Yezefski, T., Schwemm, A., Lentz, M., Hone, K. and Shankaran, V. (2018), 'Patient assistance programs: A valuable, yet imperfect, way to ease the financial toxicity of cancer care', *Seminars in Hematology* **55**(4), 185–188.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

722. N. CIURILA, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*
723. N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*
724. S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*
725. S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*
726. Y.ZHU, *On the Effects of CEO Compensation*
727. S. XIA, *Essays on Markets for CEOs and Financial Analysts*
728. I. SAKALAUSKAITE, *Essays on Malpractice in Finance*
729. M.M. GARDBERG, *Financial Integration and Global Imbalances*
730. U. THUMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change,*
731. B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*
732. G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*
733. Z.M. LI, *Econometric Analysis of High-Frequency Market Microstructure*
734. C.M. OOSTERVEEN, *Education Design Matters*
735. S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*
736. S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
737. M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
738. M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
739. R.J. DÖTTLING, *Essays in Financial Economics*
740. E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*

741. M. KUTLUAY, *The Value of (Avoiding) Malaria*
742. A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
743. B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
744. R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
745. P.M. GOLEC, *Essays in Financial Economics*
746. M. SOUVERIJN, *Incentives at Work*
747. M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*
748. P. VAN BRUGGEN, *On Measuring Preferences*
749. M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and Their Use in Econometrics*
750. S. GARCIA-MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*
751. Y. SUN, *Consumer Search and Quality*
752. I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*
753. G.W. GOY, *Modern Challenges to Monetary Policy*
754. A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*
755. J. SUN, *Tell Me How To Vote: Understanding the Role of Media in Modern Elections*
756. J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*
757. A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*
758. F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*
759. A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*
760. A. RUSU, *Essays in Public Economics*
761. M.A. COȚOFAN, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*

762. B.P.J. ANDRÉE, *Theory and Application of Dynamic Spatial Time Series Models*

763. P. PELZL, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*

764. D.M. KUNST, *Essays on Technological Change, Skill Premia and Development*

765. A.J. HUMMEL, *Tax Policy in Imperfect Labor Markets*

766. T. KLEIN, *Essays in Competition Economics*

767. M. VIGH, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*

768. YAN XU, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*

769. S. RELLSTAB, *Balancing Paid Work and Unpaid Care over the Life-Cycle*

770. Z. DENG, *Empirical Studies in Health and Development Economics*

771. L. KONG, *Identification Robust Testing in Linear Factor Models*

772. I. Neamțu, *Unintended Consequences of Post-Crisis Banking Reforms*